# Journal of Medical Internet Research

Journal Impact Factor (JIF) (2023): 5.8
Volume 18 (2016), Issue 7 ISSN 1438-8871 Editor in Chief: Gunther Eysenbach, MD, MPH

## Contents

### Viewpoints

Are Randomized Controlled Trials the (G)old Standard? From Clinical Intelligence to Prescriptive Analytics (e185)
Sven Van Poucke, Michiel Thomeer, John Heath, Milan Vukicevic. . . . . 4

Telemedical Education: Training Digital Natives in Telemedicine (e193)
Akhilesh Pathipati, Tej Azad, Kamal Jethwani. . . . . 15

### Original Papers

Innovatively Supporting Teachers' Implementation of School-Based Sex Education: Developing A Web-Based Coaching Intervention From Problem to Solution (e136)
Lisette Schutte, Marieke van den Borne, Gerjo Kok, Suzanne Meijer, Fraukje Mevissen. . . . . 19

Parent-Focused Childhood and Adolescent Overweight and Obesity eHealth Interventions: A Systematic Review and Meta-Analysis (e203)
Megan Hammersley, Rachel Jones, Anthony Okely. . . . . 39

A Mobile Ecological Momentary Assessment Tool (devilSPARC) for Nutrition and Physical Activity Behaviors in College Students: A Validation Study (e209)
Meg Bruening, Irene van Woerden, Michael Todd, Stephanie Brennhofer, Melissa Laska, Genevieve Dunton. . . . . 65

Follow #eHealth2011: Measuring the Role and Effectiveness of Online and Social Media in Increasing the Outreach of a Scientific Conference (e191)
Marcel Winandy, Patty Kostkova, Ed de Quincey, Connie St Louis, Martin Szomszor. . . . . 77

How Affiliation Disclosure and Control Over User-Generated Comments Affects Consumer Health Knowledge and Behavior: A Randomized Controlled Experiment of Pharmaceutical Direct-to-Consumer Advertising on Social Media (e189)
David DeAndrea, Megan Vendemia. . . . . 98

Estimating Influenza Outbreaks Using Both Search Engine Query Data and Social Media Data in South Korea (e177)
Hyekyung Woo, Youngtae Cho, Eunyoung Shim, Jong-Koo Lee, Chang-Gun Lee, Seong Kim. . . . . 109

Health Care Applicability of a Patient-Centric Web Portal for Patients' Medication Experience (e202)
Song Hong, Woojung Lee, Yazed AlRuthia. . . . . 120



XSL•FO
RenderX

Online Health Information-Seeking Behavior and Confidence in Filling Out Online Forms Among Latinos: A Cross-Sectional Analysis of the California Health Interview Survey, 2011-2012 (e184)
Mariaelena Gonzalez, Ashley Sanders-Jackson, Jason Emory. . . . 132

Modelling and Predicting eHealth Usage in Europe: A Multidimensional Approach From an Online Survey of 13,000 European Union Internet Users (e188)
Joan Torrent-Sellens, Ángel Díaz-Chao, Ivan Soler-Ramos, Francesc Saigí-Rubió. . . . 146

Behavioral Analysis of Visitors to a Medical Institution's Website Using Markov Chain Monte Carlo Methods (e199)
Teppei Suzuki, Yuji Tani, Katsuhiko Ogasawara. . . . 178

First-Stage Development and Validation of a Web-Based Automated Dietary Modeling Tool: Using Constraint Optimization Techniques to Streamline Food Group and Macronutrient Focused Dietary Prescriptions for Clinical Trials (e190)
Yasmine Probst, Evan Morrison, Emma Sullivan, Hoa Dam. . . . 189

An Integrated Process and Outcome Evaluation of a Web-Based Communication Tool for Patients With Malignant Lymphoma: Randomized Controlled Trial (e206)
Inge van Bruinessen, Evelyn van Weel-Baumgarten, Hans Gouw, Josée Zijlstra, Sandra van Dulmen. . . . 201

Paper-Based Survivorship Care Plans May be Less Helpful for Cancer Patients Who Search for Disease-Related Information on the Internet: Results of the Registrationsystem Oncological Gynecology (ROGY) Care Randomized Trial (e162)
Kim Nicolaije, Nicole Ezendam, Johanna Pijnenborg, Dorry Boll, Maria Vos, Roy Kruitwagen, Lonneke van de Poll-Franse. . . . 215

Sustained Use of Patient Portal Features and Improvements in Diabetes Physiological Measures (e179)
Stephanie Shimada, Jeroan Allison, Amy Rosen, Hua Feng, Thomas Houston. . . . 230

The Effect of a Freely Available Flipped Classroom Course on Health Care Worker Patient Safety Culture: A Prospective Controlled Study (e180)
Lowell Ling, Charles Gomersall, Winnie Samy, Gavin Joynt, Czarina Leung, Wai-Tat Wong, Anna Lee. . . . 243

Possible Biases of Researchers' Attitudes Toward Video Games: Publication Trends Analysis of the Medical Literature (1980–2013) (e196)
Aviv Segev, Mitchell Rovner, David Appel, Aaron Abrams, Michal Rotem, Yuval Bloch. . . . 252

Older Cancer Patients' User Experiences With Web-Based Health Information Tools: A Think-Aloud Study (e208)
Sifra Bolle, Geke Romijn, Ellen Smets, Eugene Loos, Marleen Kunneman, Julia van Weert. . . . 262

Social Annotation Valence: The Impact on Online Informed Consent Beliefs and Behavior (e197)
Martina Balestra, Orit Shaer, Johanna Okerlund, Lauren Westendorf, Madeleine Ball, Oded Nov. . . . 279

A Social Media mHealth Solution to Address the Needs of Dengue Prevention and Management in Sri Lanka (e149)
May Lwin, Santosh Vijaykumar, Vajira Rathnayake, Gentatsu Lim, Chitra Panchapakesan, Schubert Foo, Ruwan Wijayamuni, Prasad Wimalaratne, Owen Fernando. . . . 297

Linguistic and Cultural Adaptation of a Computer-Based Counseling Program (CARE+ Spanish) to Support HIV Treatment Adherence and Risk Reduction for People Living With HIV/AIDS: A Randomized Controlled Trial (e195)
Ann Kurth, Nok Chhun, Charles Cleland, Michele Crespo-Fierro, José Parés-Avila, John Lizcano, Robert Norman, Michele Shedlin, Barbara Johnston, Victoria Sharp. . . . 312



XSL•FO
RenderX

Being an Informed Consumer of Health Information and Assessment of Electronic Health Literacy in a National Sample of Internet Users: Validity and Reliability of the e-HLS Instrument (e161)
Gül Seçkin, Dale Yeatts, Susan Hughes, Cassie Hudson, Valarie Bell. . . . 328

Associations of eHealth Literacy With Health Behavior Among Adult Internet Users (e192)
Seigo Mitsutake, Ai Shibata, Kaori Ishii, Koichiro Oka. . . . 347

Which Doctor to Trust: A Recommender System for Identifying the Right Doctors (e186)
Li Guo, Bo Jin, Cuili Yao, Haoyu Yang, Degen Huang, Fei Wang. . . . 356

Influence of Intensified Supervision by Health Care Inspectorates on Online Patient Ratings of Hospitals: A Multilevel Study of More Than 43,000 Online Ratings (e198)
Rudolf Kool, Sophia Kleefstra, Ine Borghans, Femke Atsma, Tom van de Belt. . . . 367

Investigating the Potential Contribution of Patient Rating Sites to Hospital Supervision: Exploratory Results From an Interview Study in the Netherlands (e201)
Sophia Kleefstra, Linda Zandbelt, Ine Borghans, Hanneke de Haes, Rudolf Kool. . . . 375

## Reviews

Web-Based Interventions for Chronic Back Pain: A Systematic Review (e139)
Shashank Garg, Divya Garg, Tanvir Turin, M Chowdhury. . . . 52

The Personal Emergency Response System as a Technology Innovation in Primary Health Care Services: An Integrative Review (e187)
Randi Stokke. . . . 164



XSL•FO
RenderX

Viewpoint

# Are Randomized Controlled Trials the (G)old Standard? From Clinical Intelligence to Prescriptive Analytics

Sven Van Poucke[1*], MD; Michiel Thomeer[2*], MD, PhD; John Heath[3*]; Milan Vukicevic[4*], PhD

[1]Department of Anesthesiology, Critical Care, Emergency Medicine, Pain Therapy, Ziekenhuis Oost-Limburg, Genk, Belgium

[2]Department of Pneumology, Ziekenhuis Oost-Limburg, Genk, Belgium

[3]RapidMiner, Shanghai, China

[4]Faculty of Organizational Sciences, University of Belgrade, Belgrade, Serbia

[*]all authors contributed equally

**Corresponding Author:**
Sven Van Poucke, MD
Department of Anesthesiology, Critical Care, Emergency Medicine, Pain Therapy
Ziekenhuis Oost-Limburg
Schiepse Bos 6
Genk, 3600
Belgium
Phone: 32 +32475269857
Fax: 32 89327905
Email: svanpoucke@gmail.com

## Abstract

Despite the accelerating pace of scientific discovery, the current clinical research enterprise does not sufficiently address pressing clinical questions. Given the constraints on clinical trials, for a majority of clinical questions, the only relevant data available to aid in decision making are based on observation and experience. Our purpose here is 3-fold. First, we describe the classic context of medical research guided by Poppers' scientific epistemology of "falsificationism." Second, we discuss challenges and shortcomings of randomized controlled trials and present the potential of observational studies based on big data. Third, we cover several obstacles related to the use of observational (retrospective) data in clinical studies. We conclude that randomized controlled trials are not at risk for extinction, but innovations in statistics, machine learning, and big data analytics may generate a completely new ecosystem for exploration and validation.

***(J Med Internet Res 2016;18(7):e185)*** doi:10.2196/jmir.5549

**KEYWORDS**

randomized controlled trials; data mining; big data; predictive analytics; algorithm; modeling; ensemble methods

## Introduction

Despite the accelerating pace of scientific discovery, the current clinical research enterprise does not fully address daily clinical questions such as "what is the most adequate course of action for a particular patient, under these conditions, in this phase of the illness?" [1]. From a clinician's perspective, the most abundant information available for decision making is based on observation and experience [2]. With the accumulation of large amounts of health-related data, the methods for therapeutic effect quantification have been rapidly evolving and are driven by recent innovations in statistics, machine learning, and big data analytics [3]. Recent technology allows the use of (near) real-time clinical decision support tools, enabling the quantification and prioritization of unanswered clinical questions in the absence of published evidence [4].

Despite the abundance of data available, fitting data to a model to explain observations might be plausible and appear to be in agreement with clinical experiences, but the derivation of natural laws or theories cannot be justified. From an epistemological point of view (Karl Popper), science should strive to describe simple and logical theoretical systems that are testable before enabling any predictions [5]. Classically, deductive science begins with a hypothesis or theory and proceeds to derive possible conclusions and statements. With the introduction of precision medicine as an emerging approach for disease treatment and prevention, the question arises whether simple and logical theoretical systems are the only choice for predictive



XSL•FO
RenderX

analysis of complex, high-dimensional data from a multimorbidity patient population [6].

Various methods have been presented to predict future outcomes or to forecast trends using observational data [7]. Observational data research might seem attractive because of lower cost and time consumption, but it is mostly considered inferior to prospective research. In the big data and Internet of things era, "observational" data are abundant and could be considered a historical footprint, valuable for training and testing models from which performance can be quantitatively assessed using new data input [8].

The aim of this viewpoint paper is to highlight some innovations in statistics, machine learning, and big data analytics, and to confront them with the current gold standard test used in clinical trials: the randomized controlled trial (RCT). Therefore, we discuss this in three sections—challenges and shortcomings of RCTs, potential of observational studies with big data, and challenges and difficulties of observational (retrospective) data for clinical studies.

## The Challenges and Shortcomings of RCTs

RCTs were introduced in medicine more than half a century ago [9]. The trial is initiated by a null hypothesis that there is no decisive evidence that the intervention or drug being tested is superior to existing treatments. In prospective RCTs, the investigators conceive and design the trial, recruit participants, and collect baseline data, before the participants have developed any of the outcomes of interest. Individuals are selected from a population to estimate characteristics of the entire population. The intervention is randomly assigned after participants have been assessed for eligibility and recruitment, but prior to the intervention under study. When properly designed, RCTs can isolate confounding factors and allow researchers to identify causal effects between input and observed phenomena. This makes RCTs the gold standard for evidence-based medicine (EBM) [10]. The Framingham Heart Study is a historical example of a large, productive prospective cohort study [11].

In contrast, it is widely acknowledged that evidence from RCTs frequently rests on narrow patient inclusion criteria, hindering generalization to real clinical situations [12]. As such, RCTs do not ensure the translation of their results into tangible benefits to the general population [13]. Additionally, it is often unclear which assumptions are part of the hypothesis. Frequently, researchers end up with central tendencies from a group of individuals, a measure that is often not representative of an individual patient.

Limitations of RCTs or suboptimally designed RCTs are at times overlooked or ignored [14]. When RCTs lack methodological rigor, the results must be interpreted cautiously [15,16]. Furthermore, the cost and duration of RCTs may be prohibitive, delaying the acceptance of new treatment modalities [17]. The outcome of interest in RCTs should also be common; otherwise, the number of outcomes observed becomes too small for statistical meaningfulness (indistinguishable from the cases that may have arisen by chance).

Additionally, certain interventions might not be suitable to be explored by RCTs because of ethical considerations. Likewise, when an intervention becomes widespread, clinicians are unwilling to experiment with alternatives. For instance, the impact of timing of cardiopulmonary resuscitation on cerebral and myocardial functional recovery cannot be investigated with controlled trials. However, such studies can be designed using techniques such as propensity score analysis and stratification based on big data [18,19].

With the aging of the population, an increasing percentage of patients have multiple comorbidities, which are routinely excluded from RCTs. In contrast, big data from electronic medical records provide information from real-world settings [19]. Research based on these data might be more applicable to patients encountered in daily practice.

Even with a well-designed and successfully conducted RCT, many clinical questions are unanswered, because results from RCTs might not be suited to each individual patient. This problem is the main focus of personalized and precision medicine [6]. An obvious example is that over the past few decades, perioperative management has improved in safety, resulting in lower incidences of major perioperative complications (<1% to 3%), such as perioperative stroke or death. Nonetheless, even events with a 1% incidence rate would affect 2 million people each year worldwide. These devastating complications are hardly studied in RCTs, as their low incidences would require inclusion of significant numbers of patients [20]. Big data analytics might facilitate research for these rare end points, thereby potentially opening opportunities for improving clinical practice [21].

In the last two decades, EBM attempted to address the limitations of RCTs. EBM is commonly defined as "the conscientious, explicit, and judicious use of current best evidence in making decisions about the care of individual patients" [22]. The purpose of EBM is to provide a stronger scientific foundation for clinical work, so as to achieve consistency, efficiency, effectiveness, quality, and safety in medical care. The theoretical ideal of EBM, where every clinical question would be based on meta-analysis and systematic reviews of multiple RCTs, faces multiple limitations. An early example of EBM can be found in the British Thoracic Society's 1990 asthma guidelines, developed through consensus and based on a combination of randomized trials and observational studies [23].

Two decades of enthusiasm could not prevent some from arguing that the EBM movement is in crisis, for many reasons [2]: (1) the evidence-based "quality mark" has been misappropriated by vested interests, (2) the volume of evidence, especially clinical guidelines, has become unmanageable, (3) statistically significant benefits may be marginal in clinical practice, (4) inflexible rules and technology-driven prompts may produce care that is *management driven* rather than *patient centered*, and (5) evidence-based guidelines often map poorly to complex multimorbidity.

It is remarkable that recent advances in analytics are not mentioned in any "strength of evidence" rankings [24]. This closely resembles the plea from Angus Deaton, the 2015 Nobel



XSL•FO
RenderX

prize winner in economic sciences, for more modesty in what randomized trials can offer, fulminating against a one-size-fits-all mentality [25].

## Potential of Observational Studies With Big Data

The burden of chronic diseases is rapidly increasing worldwide, triggering a paradigm shift from delayed interventional to predictive, preventive, and personalized medicine [26,27]. Success stories of the big data paradigm and data mining led to broader recognition of the potential impact and benefits (both human and economic) in health care. In 2012, the worldwide amount of digital health care data was estimated to be around 500 petabytes, expected to reach 25,000 petabytes in 2020, of which approximately 80% is unstructured [28].

The explosion in data has opened a multitude of opportunities for improving health care in general by the design of data-driven models for different tasks: (1) in public health: prediction of admission rates, epidemics, hospital capacities, etc, (2) early risk prediction for mortality, hospital readmission, treatment efficacy, etc, (3) for chronic disease control: drug dosage optimization, therapeutic adherence, etc, (4) in diagnostics: decision support systems in medical imaging, etc.

Predictive modelling in a clinical context, where data are collected, a statistical model is formulated, predictions are made, and the model is validated (or revised) as additional data become available, could become the key for tailoring medical treatment to individual characteristics of each patient (precision medicine initiative [6]).

A recent report on the potential of learning health care systems suggested that the RCT is not dead, but rather that other methodologies will be required if we are to bridge the evidence gap in modern medicine [29]. Observational studies can deliver useful results quickly, at lower cost, and do not put patients at risk through experimental exposure. The development of electronic health records and rigorous outcomes measurement offers the potential to accelerate the use of observational research. This may require a paradigm shift in education and research.

Retrospective data are historically assessed by descriptive statistical analysis, resulting in clinical intelligence (Figure 1). Predictive analytics differs from clinical intelligence and business intelligence-style intelligence in its use of models—models that capture and represent hidden patterns and interactions in the data.

Clinical decisions, once exclusively guided by *experience* (wisdom generated from qualitative retrospective analysis) and *retrospective clinical intelligence* (wisdom from quantitative retrospective analysis), can now be upgraded by knowledge of predictive and prescriptive analytics, predicting future events on the individual patient level (Figure 1).

Big data is defined as high-volume, high-velocity, high-variety, and high-veracity information assets, requiring new forms of processing to enhance decision making, insight discovery, and process optimization [30]. Cutting-edge big data technologies allow for integration and scalable analytics of heterogeneous medical data. Additionally, recent computational and mathematical advances have enabled effective usage of machine learning and data mining methods for uncovering hidden relationships between different parameters and clinical outcomes [13]. This evolution is considered one of the main factors in the development of predictive, preventive, and personalized medicine. Big data might increase the relatively low ratio of screened to enrolled patients of RCTs, optimizing the generalization of results from research in routine clinical practice (external validity).

Data availability in clinical medicine can be seen as both wide (from large populations) and deep (a large amount of data per patient). Wide data allow for analytics of various trends in public health care (eg, the number of admissions per disease or hospital) and can be used in quality indicators for hospitals (eg, readmission rates), newly introduced drugs, or health campaigns. In other words, wide secondary data provide the essential raw material for key operations in health care. Plans and priorities of governmental health departments and clinical decision making based on historical disease characteristics both depend on secondary data. For example, virtually every basic-science grant application for severe sepsis research contextualizes the proposed work with national-scale epidemiology derived from administrative records [27,31]. Policy concerns about health care overuse in the intensive care unit, such as excessive end-of-life spending and unexplained geographic variation in intensive care unit use, depend on secondary data analyses [32,33]. Much of our understanding of racial or ethnic and insurance-based disparities, as well as the value of critical care, derives from secondary data analyses [34].

Directly related to the exploration of wide data, initiatives were promoted for collecting, integrating, and making publicly available these data for analyses. One of the largest open databases of this kind is the State Inpatient Databases, a US Agency for Healthcare Research and Quality Healthcare Cost and Utilization Project [35]. The State Inpatient Databases (2001–2010) include about 330 million inpatient discharges from 46 US states. These data track all hospital admissions at the individual level, and track diagnostic and procedural data based on *International Classification of Diseases, Ninth Revision, Clinical Modification* coding. Additionally, demographics and administrative data of each admission are tracked (eg, sex, age, month of admission, length of stay, and total charges in US currency). Opening up these data initiated many research efforts in health care predictive analytics as published on websites from the US National Information Center on Health Services Research and Health Care Technology and others.

However, wide data are not the best information source to generate clinically relevant research at the patient level (eg, mortality risk, evaluation of effectiveness of procedures), because these data are in most cases generated for administrative and reimbursement purposes, and are not sufficiently detailed to describe complex medical states and outcomes for a unique patient.

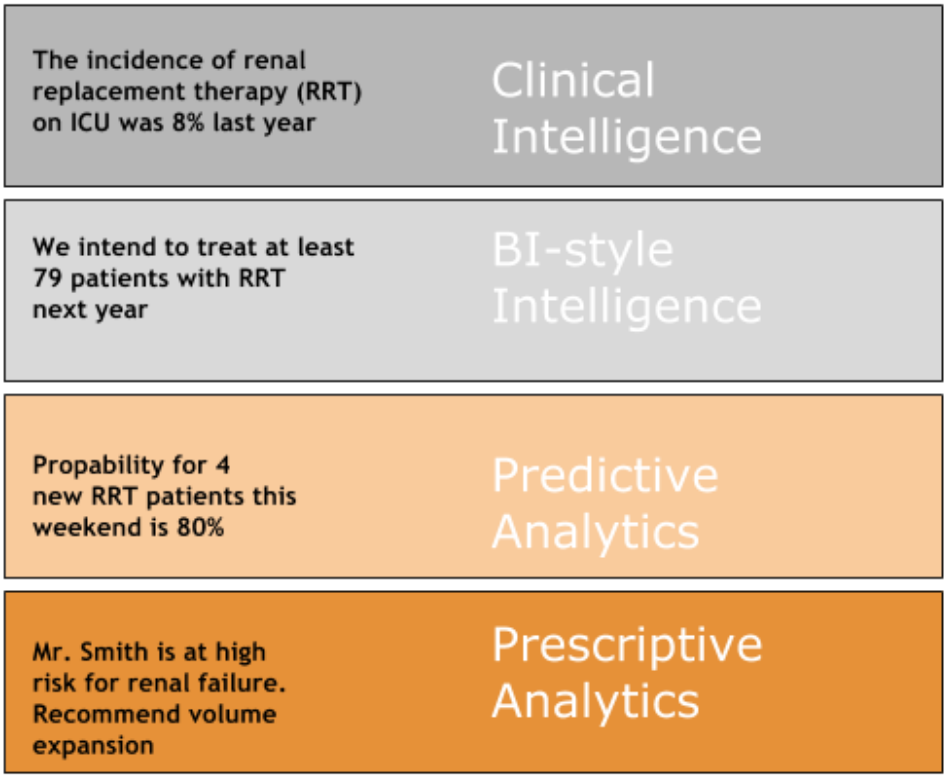


XSL•FO
RenderX

Deep data, on the other hand, provide a higher level of temporal details from each patient, on multiple scales (eg, genomics, proteomics, drugs, laboratory tests, comorbidities, symptoms). When analyzed properly, such data have the potential to provide valuable clinical insights and could change practice in fundamental ways, improving outcomes for patients [6]. A good example is the reevaluation of the use of pulmonary artery catheters, once a ubiquitous feature of the treatment of nearly every medical intensive care unit patient, but this use was reinvestigated with a clever reanalysis of a clinical trial [36].

The importance of opening deep data for analytics is recognized widely. One of the most popular and most detailed data sources available is the Multiparameter Intelligent Monitoring in Intensive Care (MIMIC) clinical database, which contains data on 58,976 intensive care unit admissions (medical, surgical, coronary care, and neonatal), for over 48,000 distinct patients admitted to Beth Israel Deaconess Medical Center (Boston, MA, USA) from 2001 to 2012 [37]. The MIMIC-III database contains highly detailed and heterogeneous data (laboratory tests, vital signs, symptoms, medical imaging, notes, waveforms, etc). The data in the MIMIC-III database are available to other researchers and there are no privacy concerns, promoting reproducibility of research. Opening this database yielded many promising research efforts [38,39].

**Figure 1.** From clinical intelligence to prescriptive analytics. BI business intelligence; ICU: intensive care unit.




## Challenges and Difficulties of Observational (Retrospective) Data for Clinical Studies

Observational studies look at medical events from some time point in the past and examine exposure to a suspected risk or treatment in relation to an outcome established at the start of the study. There are several challenges opposing the quick and easy development of predictive models with good performance, in particular for complex clinical problems.

This results in a large gap between potential and actual data usage [27,31]. Retrospective databases pose a series of methodological challenges, some of which are unique to this data source [40].

### Correlation Does Not Imply Causation

One of the major obstacles to full applicability of predictive analytics in real-life clinical practice (and distrust of observational studies) is the credibility of the evolved patterns (models). Although modelling enables quantification of correlation on large data sources, correlation in most cases does not imply causation (even with significant correlations identified). Two major fallacies have been described in this respect: *cum hoc ergo propter hoc*, Latin for "with this, therefore because of this," and *post hoc ergo propter hoc*, Latin for "after this, therefore because of this." The main cause of misleading conclusions based on identified correlations is incorrect sample or feature selection, which leads to neglect of actual confounders. Namely, retrospective studies are often conducted on large data samples, but these samples are not described with all potential confounders [41,42]. On the other hand, stratification of a population, leading to homogeneous and well-described groups (eg, the same age group, sex, history of



XSL•FO
RenderX

diseases, current health status, and vital signs), also leads to insufficient data quantities because of the complexity of medical phenomena and the large number of potential confounders. So, when a population is carefully selected, in most cases, the lack of data emerges as a problem that prevents the development of accurate and stable predictive models. In these situations, an additional problem arises in identification of real causal relationships: "the curse of dimensionality" or Hughes phenomenon [43]. The curse of dimensionality states that, with a fixed number of training samples, the predictive power reduces as the dimensionality increases, meaning that patterns identified in high-dimensional spaces may occur due to chance. Consequently, an enormous amount of data is needed to ensure that a population is well described by a given sample.

To conclude, in theory it is possible to select an adequate sample and feature space that describes well the medical phenomena that are observed, and eventually could lead to causal relationships and insights. However, finding such samples in retrospective data is very challenging, and this problem has to be addressed adequately when reporting and interpreting predictive results from retrospective studies.

### Fusion of Data Science and Domain Expertise

Even if retrospective studies are well defined (in relation to samples and features) and if the medical community is confident with models and results, successful predictive analytics and application of cutting-edge machine learning algorithms often demands substantial programming skills in different languages (eg, Python or R). This migrates modelling from the domain expert to the data scientist, often missing the necessary domain expertise, and vice versa. Additionally, data analyses are highly creative processes and there are no detailed recommendations for conducting such research. High-level steps for conducting this research is described by the cross-industry standard process for data mining, which is breaking down the life cycle of an analytics project into six phases: business understanding, data understanding, data preparation, modelling, evaluation, and deployment [44]. However, specifications of each problem prevent the development of a standardized analytics process on an operational level. This ultimately leads to the slow development, adoption, and exploitation of highly accurate predictive models, in particular in medical practice, where errors have significant consequences (both human and financial) [45]. Obviously, a close and continuous collaboration between domain experts and data scientists would solve this problem, but this is not always feasible. Many efforts have attempted to overcome this problem in recent research. One of the directions is formalization of domain knowledge through medical ontologies (eg, Disease Ontology [46], SNOMED [47], and for orofacial pain [48]) and integration with data-driven models [49,50]. This approach aims to allow for data-knowledge fusion and to reduce the need for additional specialization of domain experts in data science and vice versa. Another approach is development of visual analytics tools that enable a faster learning curve and powerful analytics that can be conducted by domain experts [45,51].

### Data Heterogeneity and Quality

In particular, deep medical data that could potentially provide meaningful clinical conclusions or new hypotheses is highly heterogeneous: laboratory tests, disease history, comorbidities (multiple diagnoses), medication prescriptions, protein interaction networks, genomic sequences, medical imaging, notes, waveforms, and so on. In addition to different data formats, the data are time stamped, temporal, context dependent, and defined over different levels of granularity. This raises the challenging problem of extracting information and meaningful patterns from all available data sources, even with cutting-edge big data technologies that allow for efficient storage and manipulation of such data and predictive methods that allow for temporal modelling of interdependent data [52]. Various ways have been proposed to address these problems, such as integrating the results of models that are built separately on homogeneous data sources, and mapping between problem (data) spaces and learning models on common data representations [53]. However, each step of these strategies loses information and propagates uncertainty, and thus the potential of big and heterogeneous data is only partially exploited. Additionally, it is essential to interpret the findings in the context of a defined patient population (generalizability). If multiple data sources were used to construct a database, it is important to emphasize whether the necessary linkages between data sources and various care sites have been carried out appropriately, taking into account differences in coding and reporting across sources and timestamping (data linkage). Retrospective data face a renewed interest with the growth of big datasets, as questions arise related to the quality of the data and the source validity. With frequently unknown quality or completeness of the recorded data, "garbage in, garbage out" (or GIGO) is commonly used to describe failures in human decision making due to faulty, incomplete [38], or imprecise data.

### Validation and Reproducibility

Even though many studies have reported cutting-edge performance in predictive modelling on biomedical data, evolved models often show unstable or unconvincing performance when applied outside of the initial experimental setting.

Some of the reasons for this are that validation measures used in experiments are misleading; that specific algorithm implementations and data are not always available; and that experimental settings are not sufficiently described and, thus, the results reported in scientific papers cannot be reproduced in other settings.

Selection of adequate validation measures is highly dependent on the nature of the data for building models. Since most of these datasets have an imbalance between the size of the positive and negative classes, classification accuracy is a meaningless performance measurement. For this reason, other evaluation criteria are used, such as the area under the receiver operating characteristic curve and the area under the precision recall curve. All of these are based on the basic notions of the numbers of true positives, false positives, true negatives, and false negatives [54,55].



XSL•FO
RenderX

Further, in order to realistically estimate model performance in the future (on unseen cases), experimental setups will need perfection and to be protected against overfitting (the situation where a model has good performance on training data but shows poor generalization performance when tested on unseen cases). As discussed before, finding the predictive model best suited to the data at hand is often based on trial and error, and assumes comparisons of multiple models with multiple parameter settings. The number of trials and the complexity of the models positively correlate with the probability of model overfitting.

This is why parameter optimization and multiple model testing should also be monitored using an alternative partition of the data (validation dataset). A common technique to validate a model is either cross-validation or bootstrap validation [56]. Cross-validation is often used to select the optimal level of complexity (maximal predictive power without overfitting).

Other methods focus on estimating heterogeneity in causal effects in experimental and observational studies, and on conducting hypothesis tests of the magnitude of the differences in treatment effects across subsets of the population. These approaches are often tailored to situations with multiple attributes of a unit relative to the number of units observed, and where the functional form of the relationship between treatment effects and the attributes of units is unknown [19].

Finally, the error rate of the model is estimated with the remaining data partition [57]. As such, the testing data represents a realistic assessment of the model's correctness when applied to new datasets. Additionally, it is utterly important to take special care when selecting data for validation and final model performance evaluation (because models could adapt and generalize well only on a subset of the data, and thus all types of data that are expected in the future have to be present in the final evaluation of the model).

When modelling is done properly, accurate predictive models have the ability to adjust and improve over time. The artificial intelligence resulting from this evolution might have the potential to measure and optimize therapeutic effect and adherence [58].

## Interpretability

In the process of building a useful representation of a system or phenomenon, interpretability (comprehensibility or ability to understand) is often recalled. This is of particular importance in the medical domain because, even with the best diagnostic assessment and highly accurate predictive models, decisions have to be made with caution and with involvement of medical experts. If models are interpretable, medical experts can put information provided by predictive models in their specific context (reducing the danger of potential confounder influence) and get better insights into the reasons for phenomena identified by predictive models. This should eventually lead to making informed decisions and taking a step toward prescriptive analytics. However, there is a clear trade-off between model complexity and model interpretability. Additionally, interpretability is in the eye of the beholder: it is hard to make some objective comparisons between predictive models. Model interpretability is also related to the number of features and the information provided by the features. The number of features is intuitively evident as an interpretability measure. The higher the dimensionality, the more complex it becomes for human beings to analyze the relative impact of features and patterns that are potentially important in making decisions. Therefore, using a reduced set of features might lead to more interpretable models (eg, through backward feature elimination, or forward feature construction). The basic principle of all predictive methods for decreasing the number of features is to extract factors from features, by mapping (transforming) the feature space to a low-dimensional space, while keeping as much of the original variance of the features as possible.

On the other hand, the contextual information provided by the features is important regardless of dimensionality. If a model is based on a limited number of features but the human interpreter considers the model to be a black box, then the model is not interpretable. Interpretability requires more thought on how the results of predictive models help in explaining an underlying phenomenon [59]. Because of this, state-of-the-art predictive algorithms, which often provide highly accurate models (eg, neural networks or support vector machines), are often not considered useful for real-life medical applications. This poses an additional challenge to making highly accurate predictive models based on less-complex and more-interpretable algorithms such as logistic regression, naive Bayes, or decision trees. Unfortunately, interpretability and accuracy are usually concurrent, and this increases the importance of feature selection and construction in predictive modelling processes.

## There Is No Free Lunch

Many predictive algorithms have been developed, but there is no evidence that any algorithm outperforms all others in every situation. Strong support for this claim is given by "no free lunch" theories [60], where researchers demonstrate that no predictive algorithm outperforms others on every dataset, but one can always find an algorithm that is optimal for a dataset. In particular, in health care predictive analytics, the consequences of no free lunch theories are posing a very challenging problem of finding the algorithm best suited to the data at hand. This is directly related to the complexity of medical phenomena, contextual dependency, data heterogeneity, high dimensionality, class imbalance, and so on. For many of these specific problems, a variety of efficient predictive methods have been developed. For example, lasso logistic regression efficiently reduces dimensionality of the initial dataset [61], while preserving or even increasing the predictive performance on unseen data. Support vector machines [62] efficiently avoid overfitting and allow incorporation of domain knowledge by kernel engineering. Neural networks and deep learning methods have the ability to fit high-dimensional data and to model spatiotemporal relations in data [63]. Further, ensemble methods [64] are used to improve the performance of individual algorithms. They have shown many advantages in dealing with a small sample size, high dimensionality, and complex data structures by exploiting the diversity among the models produced. These models can be aggregated from the same model built on different subsamples of data, from different models built on the same sample, or a combination of the previous two techniques. Some popular algorithms from this class are bagging



XSL•FO
RenderX

[65], random forest [66], boosting [67], and bootstrap aggregating.

However, all mentioned models have their own cons and there are no theoretical guarantees for a model's success in a particular application. The problem of finding the best model for a particular dataset is influenced by data preprocessing (feature selection, feature construction, etc). The objective of variable (feature) selection is 3-fold: improving the prediction performance of the predictors, providing faster and more cost-effective predictors, and providing a better understanding of the underlying process that generated the data [68]. This requires feature construction, feature ranking, multivariate feature selection, efficient search methods, and feature validity assessment methods.

### Privacy Concerns

Another problem often considered an obstacle for successful application of predictive analytics in health care is the lack of data. Data can be lacking for several reasons: rare diseases, long and expensive procedures for data collection, and confidentiality of personally sensitive information. Privacy concerns often restrict the potential of sharing the data between institutions and thus building more accurate and reliable models.

However, there are many techniques that could help in overcoming this problem and enable data sharing without fear of identifying patients without their permission. The process of privacy protection starts with traditional anonymization techniques, which map personal and hospital identity into an encrypted form. Additionally, time and duration of hospital visits are usually presented in a relative form (number of days from initial admission), while exact dates are removed. Even though these techniques can substantially reduce the risk of patient identification, the state-of-the-art predictive techniques theoretically can still identify the person based on procedures, diagnoses, and other data that cannot be encrypted if they are a basis for collaborative building and evaluation of predictive models. Thus, privacy of big data is of particular concern. These problems are often successfully solved by secure multiparty computation [69,70], where the sites cooperate to build the global prediction model without sharing the data themselves, and by randomization, where data are additionally masked by adding some controlled noise [71,72].

## Conclusion

By no means is the value of RCTs as a method for scientific experimentation questioned. We are convinced that it is far more reasonable to estimate the therapeutic effects from nonrandomized studies, based on the best available surrogate technology, than to ignore the potential richness of the available data [13]. Nonrandomized data could at least provide indicators of potential causality, ultimately triggering the initiation of randomized experiments.

A changing ecosystem of analytical methods has opened up and become available for exploration and validation. Observational studies could complement RCTs in generating hypotheses, establishing questions for future RCTs, and defining clinical conditions [73]. Drawing conclusions based on biased data or dubious analyses by threats of both external and internal validity should be monitored constantly in big data analysis to guarantee that a study measures what it set out to and that the results can be generalized from the study to the reader's patients.

As such, the data science community has a huge responsibility to eliminate the fear of using predictive modelling in health care by explaining the concepts of predictive modelling in a setting where humans are the preferred decision makers. Finally, data scientists need to create familiarity with data visualization as a channel for information sharing. Data-driven research incorporates artificial intelligence and machine learning into statistics and supports the recognition of patterns within massive datasets. Validation and interpretation of results is an essential step preceding data visualization.

### Conflicts of Interest

None declared.

### References

1. Van Poucke S, Thomeer M, Hadzic A. 2015, big data in healthcare: for whom the bell tolls? Crit Care 2015;19:171 [FREE Full text] [doi: 10.1186/s13054-015-0895-8] [Medline: 25886956]
2. Greenhalgh T, Howick J, Maskrey N. Evidence based medicine: a movement in crisis? BMJ 2014 Jun 13;348(jun13):g3725-g3725 [FREE Full text] [doi: 10.1136/bmj.g3725] [Medline: 24927763]
3. Raghupathi W, Raghupathi V. Big data analytics in healthcare: promise and potential. Health Inf Sci Syst 2014;2:3 [FREE Full text] [doi: 10.1186/2047-2501-2-3] [Medline: 25825667]
4. Longhurst C, Harrington R, Shah N. A 'green button' for using aggregate patient data at the point of care. Health Aff 2014 Jul 8;33(7):1229-1235 [FREE Full text] [doi: 10.1377/hlthaff.2014.0099]
5. Winkler R. Popper and the Omics. Frontiers Plant Sci 2016 Feb 19;7(195):1-3 [FREE Full text] [doi: 10.3389/fpls.2016.00195]
6. McCarthy M. Obama promises to defend health law and promote "precision medicine". BMJ 2015 Jan 21;350:h385 [FREE Full text] [Medline: 25609477]
7. Dinov ID. Methodological challenges and analytic opportunities for modeling and interpreting Big Healthcare Data. Gigascience 2016;5:12 [FREE Full text] [doi: 10.1186/s13742-016-0117-6] [Medline: 26918190]
8. Janke A, Overbeek D, Kocher K, Levy P. Exploring the potential of predictive analytics and big data in emergency care. Ann Emerg Med 2015 Feb;67(2):227-236 [FREE Full text] [doi: 10.1016/j.annemergmed.2015.06.024] [Medline: 26215667]



XSL•FO
RenderX

9. [No authors listed]. Streptomycin treatment of pulmonary tuberculosis. Br Med J 1948 Oct 30;2(4582):769-782 [FREE Full text] [Medline: 18890300]
10. Schulz K, Altman D, Moher D. CONSORT 2010 statement: updated guidelines for reporting parallel group randomized trials. Ann Intern Med 2010 Jun 1;152(11):726-732. [doi: 10.7326/0003-4819-152-11-201006010-00232] [Medline: 20335313]
11. Mahmood S, Levy D, Vasan R, Wang T. The Framingham Heart Study and the epidemiology of cardiovascular disease: a historical perspective. Lancet 2014 Mar 15;383(9921):999-1008 [FREE Full text] [doi: 10.1016/S0140-6736(13)61752-3] [Medline: 24084292]
12. Britton A, McKee M, Black N, McPherson K, Sanderson C, Bain C. Threats to applicability of randomised trials: exclusions and selective participation. J Health Serv Res Policy 1999 Apr;4(2):112-121. [doi: 10.1177/135581969900400210] [Medline: 10387403]
13. Booth C, Tannock I. Randomised controlled trials and population-based observational research: partners in the evolution of medical evidence. Br J Cancer 2014 Feb 4;110(3):551-555 [FREE Full text] [doi: 10.1038/bjc.2013.725] [Medline: 24495873]
14. Jager LR, Leek JT. An estimate of the science-wise false discovery rate and application to the top medical literature. Biostatistics 2014 Jan;15(1):1-12 [FREE Full text] [doi: 10.1093/biostatistics/kxt007] [Medline: 24068246]
15. Roberts L, Ahmed I, Hall S, Davison A. Intercessory prayer for the alleviation of ill health. Cochrane Database Syst Rev 2009 Apr 15(2):CD000368 [FREE Full text] [doi: 10.1002/14651858.CD000368.pub3] [Medline: 19370557]
16. Smith GCS, Pell JP. Parachute use to prevent death and major trauma related to gravitational challenge: systematic review of randomised controlled trials. BMJ 2003 Dec 20;327(7429):1459-1461 [FREE Full text] [doi: 10.1136/bmj.327.7429.1459] [Medline: 14684649]
17. Eapen Z, Vavalle J, Granger C, Harrington R, Peterson E, Califf R. Rescuing clinical trials in the United States and beyond: a call for action. Am Heart J 2013 Jun;165(6):837-847. [doi: 10.1016/j.ahj.2013.02.003] [Medline: 23708153]
18. Zigler CM, Dominici F. Uncertainty in propensity score estimation: Bayesian methods for variable selection and model averaged causal effects. J Am Stat Assoc 2014 Jan 1;109(505):95-107 [FREE Full text] [doi: 10.1080/01621459.2013.869498] [Medline: 24696528]
19. Zhang Z. Big data and clinical research: focusing on the area of critical care medicine in mainland China. Quant Imaging Med Surg 2014 Aug 27;4(5):426-429 [FREE Full text]
20. Mashour G, Shanks A, Kheterpal S. Perioperative stroke and associated mortality after noncardiac, nonneurologic surgery. Anesthesiology 2011 Jun;114(6):1289-1296. [doi: 10.1097/ALN.0b013e318216e7f4] [Medline: 21478735]
21. Ratliff JK, Balise R, Veeravagu A, Cole TS, Cheng I, Olshen RA, et al. Predicting occurrence of spine surgery complications using “big data” modeling of an administrative claims database. J Bone Joint Surg Am 2016 May 18;98(10):824-834. [doi: 10.2106/JBJS.15.00301] [Medline: 27194492]
22. Sackett DL, Rosenberg WMC, Gray JAM, Haynes RB, Richardson WS. Evidence based medicine: what it is and what it isn't. BMJ 1996 Jan 13;312(7023):71-72 [FREE Full text] [Medline: 8555924]
23. British Thoracic Society, Research Unit of the Royal College of Physicians of London, King's Fund Centre, National Asthma Campaign. Guidelines for management of asthma in adults: I: chronic persistent asthma. BMJ 1990;301(6753):651-653 [FREE Full text] [Medline: 29709062]
24. Chung KC, Swanson JA, Schmitz D, Sullivan D, Rohrich RJ. Introducing evidence-based medicine to plastic and reconstructive surgery. Plast Reconstr Surg 2009 Apr;123(4):1385-1389 [FREE Full text]
25. McGreevey W. The Great Escape: Health, Wealth, and the Origins of Inequality. Princeton, NJ: Princeton University Press; Dec 05, 2013:717-721.
26. Hood L, Friend S. Predictive, personalized, preventive, participatory (P4) cancer medicine. Nat Rev Clin Oncol 2011 Mar;8(3):184-187. [doi: 10.1038/nrclinonc.2010.227] [Medline: 21364692]
27. Golubnitschaja O, Kinkorova J, Costigliola V. Predictive, preventive and personalised medicine as the hardcore of 'Horizon 2020': EPMA position paper. EPMA J 2014;5(1):6 [FREE Full text] [doi: 10.1186/1878-5085-5-6] [Medline: 24708704]
28. Hersh W, Jacko JA, Greenes R, Tan J, Janies D, Embi PJ, et al. Health-care hit or miss? Nature 2011 Feb 17;470(7334):327-329. [doi: 10.1038/470327a] [Medline: 21331020]
29. Foley T, Fairmichael F. Improving the Underlying Data Focus Group.: Learning Healthcare Systems; 2015. URL: http://www.learninghealthcareproject.org/section/evidence/12/69/improving-the-underlying-data-focus-group [accessed 2016-06-29] [WebCite Cache ID 6id44BElY]
30. De Mauro A, Greco M, Grimaldi M. A formal definition of big data based on its essential features. Library Rev 2016 Apr 04;65(3):122-135. [doi: 10.1108/LR-06-2015-0061]
31. Hripcsak G, Duke J, Shah N, Reich C, Huser V, Schuemie MS, et al. Observational Health Data Sciences and Informatics (OHDSI): opportunities for observational researchers. Stud Health Technol Inform 2015;216:574-578 [FREE Full text] [Medline: 26262116]
32. Angus D, Linde-Zwirble W, Lidicker J, Clermont G, Carcillo J, Pinsky MR. Epidemiology of severe sepsis in the United States: analysis of incidence, outcome, and associated costs of care. Crit Care Med 2001 Jul;29(7):1303-1310. [Medline: 11445675]




XSL•FO
RenderX

33. Seymour C, Iwashyna T, Ehlenbach W, Wunsch H, Cooke CR. Hospital-level variation in the use of intensive care. Health Serv Res 2012 Oct;47(5):2060-2080 [FREE Full text] [doi: 10.1111/j.1475-6773.2012.01402.x] [Medline: 22985033]
34. Erickson S, Vasilevskis E, Kuzniewicz M, Cason BA, Lane RK, Dean ML, et al. The effect of race and ethnicity on outcomes among patients in the intensive care unit: a comprehensive study involving socioeconomic status and resuscitation preferences. Crit Care Med 2011 Mar;39(3):429-435 [FREE Full text] [doi: 10.1097/CCM.0b013e318206b3af] [Medline: 21187746]
35. Smith MW, Friedman B, Karaca Z, Wong HS. Predicting inpatient hospital payments in the United States: a retrospective analysis. BMC Health Serv Res 2015 Sep;15:372 [FREE Full text] [doi: 10.1186/s12913-015-1040-8] [Medline: 26358055]
36. Sandham JD, Hull RD, Brant RF, Knox L, Pineo GF, Doig CJ, Canadian Critical Care Clinical Trials Group. A randomized, controlled trial of the use of pulmonary-artery catheters in high-risk surgical patients. N Engl J Med 2003 Jan 2;348(1):5-14. [doi: 10.1056/NEJMoa021108] [Medline: 12510037]
37. Scott DJ, Lee J, Silva I, Park S, Moody GB, Celi LA, et al. Accessing the public MIMIC-II intensive care relational database for clinical research. BMC Med Inform Decis Mak 2013;13:9 [FREE Full text] [doi: 10.1186/1472-6947-13-9] [Medline: 23302652]
38. Ghassemi M, Richter S, Eche I, Chen T, Danziger J, Celi L. A data-driven approach to optimized medication dosing: a focus on heparin. Intensive Care Med 2014 Sep;40(9):1332-1339 [FREE Full text] [doi: 10.1007/s00134-014-3406-5] [Medline: 25091788]
39. Lehman L, Adams R, Mayaud L, Moody G, Malhotra A, Mark R, et al. A physiological time series dynamics-based approach to patient monitoring and outcome prediction. IEEE J Biomed Health Inform 2015 May;19(3):1068-1076 [FREE Full text] [doi: 10.1109/JBHI.2014.2330827] [Medline: 25014976]
40. Motheral B, Brooks J, Clark M, Crown W, Davey P, Hutchins D, et al. A checklist for retrospective database studies - Report of the ISPOR task force on retrospective databases. Value in Health 2003;6:90-97 [FREE Full text]
41. Sedgwick P. A comparison of sampling error and standard error. BMJ 2015;351:h3577 [FREE Full text] [doi: 10.1136/bmj.h3577]
42. Rosenbaum PR. Heterogeneity and causality. Am Stat 2005 May;59(2):147-152. [doi: 10.1198/000313005X42831]
43. Hughes G. On the mean accuracy of statistical pattern recognizers. IEEE Trans Inform Theory 1968 Jan;14(1):55-63. [doi: 10.1109/TIT.1968.1054102]
44. Shearer C. The CRISP-DM model: the new blueprint for data mining. J Data Warehousing 2000;5(4):22 [FREE Full text]
45. Van Poucke S, Zhang Z, Schmitz M, Vukicevic M, Laenen MV, Celi LA, et al. Scalable predictive analysis in critically ill patients using a visual open data analysis platform. PLoS One 2016 Jan 5;11(1):e0145791 [FREE Full text] [doi: 10.1371/journal.pone.0145791] [Medline: 26731286]
46. Kibbe WA, Arze C, Felix V, Mitraka E, Bolton E, Fu G, et al. Disease Ontology 2015 update: an expanded and updated database of human diseases for linking biomedical knowledge through disease data. Nucleic Acids Res 2015 Jan;43(Database issue):D1071-D1078 [FREE Full text] [doi: 10.1093/nar/gku1011] [Medline: 25348409]
47. Lee DH, Lau FY, Quan H. A method for encoding clinical datasets with SNOMED CT. BMC Med Inform Decis Mak 2010;10:53 [FREE Full text] [doi: 10.1186/1472-6947-10-53] [Medline: 20849611]
48. Ceusters W, Michelotti A, Raphael KG, Durham J, Ohrbach R. Perspectives on next steps in classification of oro-facial pain - part 1: role of ontology. J Oral Rehabil 2015 Dec;42(12):926-941. [doi: 10.1111/joor.12336] [Medline: 26212927]
49. Zhang GQ, Sahoo SS, Lhatoo SD. From classification to epilepsy ontology and informatics. Epilepsia 2012 Jul;53 Suppl 2:28-32 [FREE Full text] [doi: 10.1111/j.1528-1167.2012.03556.x] [Medline: 22765502]
50. Malladi V, Erickson D, Podduturi N, Rowe L, Chan E, Davidson J, et al. Ontology application and use at the ENCODE DCC. Database (Oxford) 2015;2015:1-11 [FREE Full text] [doi: 10.1093/database/bav010] [Medline: 25776021]
51. Oldham P, Hall S, Burton G. Synthetic biology: mapping the scientific landscape. PLoS One 2012;7(4):e34368 [FREE Full text] [doi: 10.1371/journal.pone.0034368] [Medline: 22539946]
52. Wolstencroft K, Owen S, Krebs O, Nguyen Q, Stanford N, Golebiewski M, et al. SEEK: a systems biology data and model management platform. BMC Syst Biol 2015;9:a [FREE Full text] [doi: 10.1186/s12918-015-0174-y]
53. Schütz C, Neumayr B, Schrefl M. Integration and reuse of heterogeneous information: hetero-homogeneous data warehouse modeling in the common warehouse metamodel. Inth Americas Conference on Information Systems 2012, AMCIS 2012.Vol 5. 2012 Presented at: 18th Americas Conference on Information Systems; January 2012; Seattle p. 18 URL: http://www.scopus.com/inward/record.url?eid=2-s2.0-84877877956&partnerID=40&md5=72a28a4e162af460481c9a4e845f98c0
54. Ozenne B, Subtil F, Maucort-Boulch D. The precision-recall curve overcame the optimism of the receiver operating characteristic curve in rare diseases. J Clin Epidemiol 2015 Aug;68(8):855-859. [doi: 10.1016/j.jclinepi.2015.02.010] [Medline: 25881487]
55. Kamath U, Jong KD, Shehu A. Effective automated feature construction and selection for classification of biological sequences. PLoS One 2014;9(7):e99982 [FREE Full text] [doi: 10.1371/journal.pone.0099982]
56. Borra S, Di Ciaccio A. Measuring the prediction error: a comparison of cross-validation, bootstrap and covariance penalty methods. Computational Stat Data Analysis 2010 Dec;54(12):2976-2989. [doi: 10.1016/j.csda.2010.03.004]
57. Williams G. Data mining with Rattle and R. J Appl Stat 2011:374.




XSL•FO
RenderX

58. Lo-Ciganic W, Donohue JM, Thorpe JM, Perera S, Thorpe CT, Marcum ZA, et al. Using machine learning to examine medication adherence thresholds and risk of hospitalization. Med Care 2015 Aug;53(8):720-728. [doi: 10.1097/MLR.0000000000000394] [Medline: 26147866]
59. Féraud R, Clérot F. A methodology to explain neural network classification. Neural Networks 2002 Mar;15(2):237-246. [Medline: 12022511]
60. Wolpert D. The lack of a priori distinctions between learning algorithms. Neural Computation 1996 Oct;8(7):1341-1390. [doi: 10.1162/neco.1996.8.7.1341]
61. Zhang Y, Biswas S. An improved version of logistic Bayesian LASSO for detecting rare haplotype-environment interactions with application to lung cancer. Cancer Inform 2015;14(Suppl 2):11-16 [FREE Full text] [doi: 10.4137/CIN.S17290] [Medline: 25733797]
62. Kong Y, Qu D, Chen X, Gong Y, Yan A. Self-organizing map (SOM) and support vector machine (SVM) models for the prediction of human epidermal growth factor receptor (EGFR/ ErbB-1) inhibitors. Comb Chem High Throughput Screen 2016;19(5):400-411. [Medline: 27074760]
63. Ghesu F, Krubasik E, Georgescu B, Singh V, Zheng Y, Hornegger J, et al. Marginal space deep learning: eficient architecture for volumetric image parsing. IEEE Trans Med Imaging 2016 Mar 7;35(5):1217-1228. [doi: 10.1109/TMI.2016.2538802] [Medline: 27046846]
64. Yang P, Hwa Yang Y, Zhou B, Zomaya A. A review of ensemble methods in bioinformatics. CBIO 2010 Dec 01;5(4):296-308 [FREE Full text] [doi: 10.2174/157489310794072508]
65. Wu M, Liao L, Luo X, Ye X, Yao Y, Chen P, et al. Analysis and classification of stride patterns associated with children development using gait signal dynamics parameters and ensemble learning algorithms. Biomed Res Int 2016;2016:9246280 [FREE Full text] [doi: 10.1155/2016/9246280] [Medline: 27034952]
66. Fu G, Ding Y, Seal A, Chen B, Sun Y, Bolton E. Predicting drug target interactions using meta-path-based semantic network analysis. BMC Bioinformatics 2016 Apr;17:160 [FREE Full text] [doi: 10.1186/s12859-016-1005-x] [Medline: 27071755]
67. Kim S, Scalzo F, Telesca D, Hu X. Ensemble of sparse classifiers for high-dimensional biological data. Int J Data Min Bioinform 2015;12(2):167-183. [Medline: 26510301]
68. Guyon I, Elisseeff A. An introduction to variable and feature selection. J Machine Learning Res 2003 Mar 03;3:1157-1182.
69. Kantarcioglu M, Nix R, Vaidya J. An efficient approximate protocol for privacy-preserving association rule mining. In: Theeramunkong T, editor. Advances in Knowledge Discovery and Data Mining: 13th Pacific-Asia Conference, PAKDD 2009 Bangkok, Thailand, April 27-30, 2009 Proceedings (Lecture Notes in Artificial Intelligence). Berlin: Springer; 2009:515-524.
70. Stiglic G, Wang F, Davey A, Obradovic Z. Pediatric readmission classification using stacked regularized logistic regression models. In: AMIA Annual Symposium Proceedings. 2014 Presented at: American Medical Informatics Association; November 15-19, 2014; Washington p. 1072.
71. Mathew G, Obradovic Z. Distributed privacy-preserving decision support system for highly imbalanced clinical data. ACM Trans Manage Inf Syst 2013 Oct 01;4(3):1-15. [doi: 10.1145/2517310]
72. Radovanovic S, Stiglic G, Delibasic B, Van Poucke S, Obradovic Z. A data and knowledge driven randomization technique for privacy-preserving data enrichment in hospital readmission prediction. In: 5th Workshop on Data Mining for Medicine and Healthcare. 2016 May Presented at: SIAM International Conference on Data Mining (SDM); May 2016; Miami, FL.
73. Song JW, Chung KC. Observational studies: cohort and case-control studies. Plast Reconstr Surg 2010 Dec;126(6):2234-2242 [FREE Full text] [doi: 10.1097/PRS.0b013e3181f44abc] [Medline: 20697313]


## Abbreviations

**EBM:** evidence-based medicine
**MIMIC:** Multiparameter Intelligent Monitoring in Intensive Care
**RCT:** randomized controlled trial


*Edited by G Eysenbach; submitted 21.01.16; peer-reviewed by L Celi, I Brooks, J Luo; comments to author 06.04.16; revised version received 01.06.16; accepted 21.06.16; published 06.07.16.*

*Please cite as:*
*Van Poucke S, Thomeer M, Heath J, Vukicevic M*
*Are Randomized Controlled Trials the (G)old Standard? From Clinical Intelligence to Prescriptive Analytics*
*J Med Internet Res 2016;18(7):e185*
*URL: http://www.jmir.org/2016/7/e185/*
*doi:10.2196/jmir.5549*
*PMID:27383622*




XSL•FO
RenderX

©Sven Van Poucke, Michiel Thomeer, John Heath, Milan Vukicevic. Originally published in the Journal of Medical Internet Research (http://www.jmir.org), 06.07.2016. This is an open-access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/2.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on http://www.jmir.org/, as well as this copyright and license information must be included.



XSL•FO
RenderX

Viewpoint

# Telemedical Education: Training Digital Natives in Telemedicine

Akhilesh S Pathipati[1*], BA; Tej D Azad[1*], BA; Kamal Jethwani[2,3], MPH, MD

[1]Stanford University School of Medicine, Stanford, CA, United States
[2]Partners HealthCare Connected Health, Boston, MA, United States
[3]Harvard Medical School, Boston, MA, United States
[*]these authors contributed equally

**Corresponding Author:**
Kamal Jethwani, MPH, MD
Partners HealthCare Connected Health
25 New Chardon St.
Boston, MA, 02114
United States
Phone: 1 6177242724
Fax: 1 6177242724
Email: kjethwani@partners.org

## Abstract

Telemedicine plays an important role in the delivery of medical care, and will become increasingly prominent going forward. Current medical students are among the first generation of "digital natives" who are well versed in the incorporation of technology into social interaction. These students are well positioned to apply advances in communications to patient care. Even so, providers require training to effectively leverage these opportunities. Therefore, we recommend introducing telemedicine training into medical school curricula and propose a model for incorporation.

***(J Med Internet Res 2016;18(7):e193)*** doi:10.2196/jmir.5534

**KEYWORDS**

telemedicine; medical education; medical school; curriculum reform

## Telemedicine Overview

Telemedicine refers to the remote delivery of medical care. Doctors have communicated over distance with one another and with patients ever since the advent of the earliest communications tools. However, recent technological advances and a changing health care landscape have transformed telemedicine from a novelty into a booming industry.

Although estimates vary, analysts project the telemedicine market to be US $20-$30 billion by 2020, with more than 100 million e-visits happening annually [1]. Nearly half of all hospitals in the United States have active telemedicine programs and are employing increasingly sophisticated tools [2]. Traditional models focus on telephone, email, and videoconferencing to care for minor conditions. These modalities remain relevant, but the field has rapidly added capabilities and indications [3]. Telemedicine includes diagnostics, treatment, monitoring, consultation, and education among other domains.

Telemedicine has become a fundamental piece of American health care delivery because it helps address issues of both health care costs and access. Moving forward, digital health capabilities will only continue to grow. In order to most effectively leverage these tools, we must ensure providers use them effectively and appropriately. Today's medical trainees are well versed in technology, but the practice of telemedicine is not necessarily intuitive. Therefore, we advocate the introduction of telemedicine training into medical schools.

## Ensuring High-Quality Care

Few among us would claim the ability to conduct an engaging conversation as a guarantee of prowess in eliciting a comprehensive patient history. Similarly, we should refrain from assuming that digital native physicians will deliver high-quality telemedical care without formal and systematic training. Current research suggests that telemedicine has a great deal of promise, but successful studies are typically carried out in academic medical centers by a limited number of well-trained doctors [4]. Other studies have shown that telemedicine can lead to mixed-quality care. For instance, Mehrotra et al [5] found that e-visits had roughly the same treatment outcomes as in-person visits for sinusitis and urinary tract infections, but e-visits had higher rates of antibiotic prescription. Schoenfeld



XSL•FO
RenderX

et al [6] found considerable variation in the quality of care provided by commercial telemedicine companies. As more patients are seen remotely and indications for telemedicine become more complex, we need to train physicians to offer digital care on par with in-person consultation.

Medical education must recognize the intrinsic differences between the practice of traditional medicine and that of telemedicine. For instance, it is difficult to remotely carry out a physical exam, which fundamentally changes the diagnostic process. Technological limitations may cause marked variation in data quality between clinic and remote visits. A patient's self-reported blood pressure from home may differ from that measured by a nurse in clinic. Providers need to be able to judge those differences. Telemedicine has its limitations in other dimensions as well. Pain management is difficult to gauge from afar. Complex diagnoses and the initial phases of patient education may be better done in person. The nature of the doctor-patient relationship is different. The list goes on.

Given these limitations, practitioners must be able to determine when telemedicine is appropriate and how to optimally process information when they see patients remotely. They must also understand how to navigate the many medicolegal issues that remain in telemedicine, including the role of Health Insurance Portability and Accountability Act regulations, restrictions due to licensing laws, and issues regarding malpractice. Telemedicine is a rapidly evolving field with many stakeholders and murky regulation; providers must learn how to interact with such a system.

## The Role of Telemedical Training

Formal training is the best way to teach providers how to approach the challenges and opportunities inherent in telemedicine. We propose that this training should begin in medical school.

Today's medical trainees are the first generation of digital natives—individuals who grew up surrounded by digital technology and are therefore comfortable processing information in an electronic world. This fact is not enough to guarantee high-quality telemedical care. Formal training can extend and amplify the impact that telemedicine brings to health care. Consider the analogy of a young athlete: the first time a tennis racket is in her hand, it is an extension of her right arm, her forehand develops easily, and she demonstrates the footwork and court instincts of a player twice her age. She's a natural. However, the distance between her innate ability and a professional career, let alone a legacy of greatness akin to that of Serena Williams, is vast. Dedicated training and repeated practice will determine whether she competes at the game's highest level. Current medical students' inherent comfort with technology should be nurtured through structured training. Without this, providers will be ill prepared to take advantage of new innovations in telemedicine.

With this in mind, we propose incorporating telemedical training into the standardized medical school curriculum. We have an opportunity to translate students' familiarity with technology into superior medical care. Creating a formal training program will allow students to directly compare and contrast telemedicine with traditional medicine, recognize when to use it, and learn best practices. Placing the training program in medical schools would ensure that all new doctors have that ability. To ensure high-quality telemedical care, we must train students to practice telemedicine with the same level of skill they demonstrate delivering traditional care.

## A Model for Incorporation

Although creating any new medical education program can seem daunting, we believe telemedicine education can be readily incorporated. Nascent efforts that expose medical trainees to telemedicine have already proven to be successful. For instance, dermatology residents and medical students on a dermatology rotation at the Denver Department of Veterans Affairs Medical Center participated in teledermatology consultations with faculty oversight [7]. Trainees reported that it was a valuable educational tool, both in terms of developing medical knowledge as well as improving their ability to provide patient care. Pilot programs at other institutions have also begun to evaluate the role of telemedicine in medical education [8-9].

To date, telemedicine training has been limited to small research settings, such as those described previously. We believe it should become a more prominent part of the medical school curriculum moving forward. Two of the authors (ASP and TDA) are students at the Stanford University School of Medicine. As such, we will use Stanford's curriculum as a theoretical model for how telemedicine education can be built into medical training.

The first two years at Stanford are the "preclinical" years, during which students take classes in the basic sciences, as well as a clinical skills class known as "Practice of Medicine" (POM). POM takes place during two 4-hour sessions each week (8 hours/week total) throughout the first two years. Students attend lectures on the process of clinical reasoning, learn how to do a history and physical exam with standardized patients (ie, actors who are pretending to be patients), work through patient cases in groups, and spend several afternoons in Stanford Hospital honing those skills with real patients. The third and fourth years of medical school are the "clinical" years, during which students rotate through various different specialties and participate in patient care.

Telemedicine training may be incorporated into both phases of medical school. During the preclinical years, one POM session every 2 months could be modified such that students must interact with patients electronically rather than in person. The clinical reasoning lectures that take place before these sessions should highlight the salient differences between electronic and traditional encounters, such as how to conduct an encounter without the physical exam, overviews of available health technologies, etc. Further research should be conducted on how to conduct a safe and effective virtual exam [10], which can then be translated into best practices.

The process of setting up a telemedicine experience during rotations is even more intuitive. Many specialties are amenable to telemedicine, including radiology, dermatology, and primary



XSL•FO
RenderX

care, among others. Students rotating in these specialties should be required to complete 10 to 20 hours on "digital call," during which they would participate in electronic encounters with faculty supervision, learn about remote monitoring tools, and develop the background necessary to be an effective provider in the future. Schools may also consider the idea of a "digital health rotation," in which students would spend 2 to 4 weeks learning how new tools can be applied in practice across fields. Granted, not all medical schools currently have the technological infrastructure in place to offer a digital call experience, but we expect those capabilities to develop as telemedicine continues to grow.

Although these suggestions are based on Stanford's curriculum, nearly every medical school in the country has a clinical skills class during the preclinical years, and clinical rotations during the final two years of medical school. Therefore, we expect the model to be generalizable to most medical schools in the United States. Further research should be conducted on specific skills and techniques that go into a safe and effective virtual encounter.

American health care is in the midst of a transformation, and telemedicine will be a cornerstone of the result. Proper training will allow us to maximize its potential.

## Conflicts of Interest

None declared.

## References


1. Deloitte. 2014. eVisits: the 21st century housecall URL: http://www2.deloitte.com/qa/en/pages/technology-media-and-telecommunications/articles/2014predictions-eVisits.html [accessed 2016-07-08] [WebCite Cache ID 6iqlllYBy]
2. American Telemedicine Association. 2014. Telemedicine frequently asked questions (FAQs) URL: http://www.americantelemed.org/about-telemedicine/faqs [accessed 2016-07-08] [WebCite Cache ID 6iqlovg77]
3. Di Cerbo A, Morales-Medina JC, Palmieri B, Iannitti T. Narrative review of telemedicine consultation in medical practice. Patient Prefer Adherence 2015;9:65-75 [FREE Full text] [doi: 10.2147/PPA.S61617] [Medline: 25609928]
4. Flodgren G, Rachas A, Farmer AJ, Inzitari M, Shepperd S. Interactive telemedicine: effects on professional practice and health care outcomes. Cochrane Database Syst Rev 2015;9:CD002098. [doi: 10.1002/14651858.CD002098.pub2] [Medline: 26343551]
5. Mehrotra A, Paone S, Martich GD, Albert SM, Shevchik GJ. A comparison of care at e-visits and physician office visits for sinusitis and urinary tract infection. JAMA Intern Med 2013 Jan 14;173(1):72-74 [FREE Full text] [doi: 10.1001/2013.jamainternmed.305] [Medline: 23403816]
6. Schoenfeld AJ, Davies JM, Marafino BJ, Dean M, DeJong C, Bardach NS, et al. Variation in quality of urgent health care provided during commercial virtual visits. JAMA Intern Med 2016 Apr 4;176(5):635-642. [doi: 10.1001/jamainternmed.2015.8248] [Medline: 27042813]
7. Boyers LN, Schultz A, Baceviciene R, Blaney S, Marvi N, Dellavalle RP, et al. Teledermatology as an educational tool for teaching dermatology to residents and medical students. Telemed J E Health 2015 Apr;21(4):312-314 [FREE Full text] [doi: 10.1089/tmj.2014.0101] [Medline: 25635528]
8. Shaikh N, Lehmann CU, Kaleida PH, Cohen BA. Efficacy and feasibility of teledermatology for paediatric medical education. J Telemed Telecare 2008;14(4):204-207. [doi: 10.1258/jtt.2008.071108] [Medline: 18534955]
9. Williams CM, Kedar I, Smith L, Brandling-Bennett HA, Lugn N, Kvedar JC. Teledermatology education for internal medicine residents. J Am Acad Dermatol 2005 Jun;52(6):1098-1099. [doi: 10.1016/j.jaad.2005.01.111] [Medline: 15928638]
10. DeJong C, Lucey CR, Dudley RA. Incorporating a new technology while doing no harm, virtually. JAMA 2015 Dec 8;314(22):2351-2352. [doi: 10.1001/jama.2015.13572] [Medline: 26647252]


## Abbreviations

**POM:** Practice of Medicine


*Edited by G Eysenbach; submitted 27.01.16; peer-reviewed by R Dudley, E Afari-kumah; comments to author 03.04.16; revised version received 05.04.16; accepted 24.04.16; published 12.07.16.*

*Please cite as:*
*Pathipati AS, Azad TD, Jethwani K*
*Telemedical Education: Training Digital Natives in Telemedicine*
*J Med Internet Res 2016;18(7):e193*
*URL: http://www.jmir.org/2016/7/e193/*
*doi:10.2196/jmir.5534*
*PMID:27405323*




XSL•FO
RenderX

©Akhilesh S Pathipati, Tej D Azad, Kamal Jethwani. Originally published in the Journal of Medical Internet Research (http://www.jmir.org), 12.07.2016. This is an open-access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/2.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on http://www.jmir.org/, as well as this copyright and license information must be included.



XSL•FO
RenderX

Original Paper

# Innovatively Supporting Teachers' Implementation of School-Based Sex Education: Developing A Web-Based Coaching Intervention From Problem to Solution

Lisette Schutte[1,2*], MPH; Marieke van den Borne[2*], MPH; Gerjo Kok[1], PhD; Suzanne Meijer[2], MPH; Fraukje EF Mevissen[1], PhD

[1]Maastricht University, Department of Work and Social Psychology, Faculty of Psychology and Neuroscience, Maastricht, Netherlands

[2]STI AIDS Netherlands, Department of Youth, Amsterdam, Netherlands

[*]these authors contributed equally

**Corresponding Author:**
Lisette Schutte, MPH
Maastricht University
Department of Work and Social Psychology
Faculty of Psychology and Neuroscience
PO Box 616
6200 MD
Maastricht,
Netherlands
Phone: 31 4338 81 908
Fax: 31 433671032
Email: lisette.schutte@maastrichtuniversity.nl

## Abstract

**Background:** Full program implementation is crucial for effectiveness but is often overlooked or insufficiently considered during development of behavioral change interventions. For school-based health promotion programs, teachers are key players in program implementation, but teacher support in this phase is mostly limited to technical support and information. To ensure optimal implementation of the Dutch school-based sexual health program Long Live Love, a Web-based coaching website was developed to support teachers in completeness and fidelity of program implementation.

**Objective:** The aim of this paper is to provide insight into the process of systematic development of a Web-based coaching intervention to support teachers in their implementation of a school-based sexual health program.

**Methods:** The intervention mapping (IM) protocol was applied for the development of a theory- and evidence-based intervention. The IM process begins with (1) a needs assessment, followed by (2) the formulation of change objectives, (3) the selection of theory-based intervention methods and practical applications that take the parameters for effectiveness into consideration, (4) integration of practical applications into an organized program, (5) planning for adoption, implementation, and sustainability of the program, and finally, (6) generating an evaluation plan to measure program effectiveness.

**Results:** Teacher's implementation behavior was characterized by inconsistently selecting parts of the program and not delivering (all) lessons as intended by program developers. Teachers, however, did not perceive this behavior as problematic, revealing the discrepancy between teacher's actual and perceived need for support in delivering Long Live Love lessons with completeness and fidelity. Teachers did, however, acknowledge different difficulties they encountered which could potentially negatively influence the quality of implementation. With the IM protocol, this Web-based coaching intervention was developed based on a concept of unobtrusive coaching, by and for teachers, to bring about change in teachers' implementation behavior.

**Conclusions:** This paper provides an example of a Web-based intervention to bring about behavioral change in a target group of intermediaries who lack intrinsic motivation for coaching and who's perceptions differ from their actual problematic behavior. The IM protocol is a useful tool for guiding the scientific development of interventions and making them compatible with the needs and preferences of the target group.

***(J Med Internet Res 2016;18(7):e136)*** doi:10.2196/jmir.5058



XSL•FO
RenderX

**KEYWORDS**

implementation; web-based coaching; intervention mapping; sexual education; unobtrusive; fidelity; teachers

## Introduction

Schools provide the ideal setting to reach youngsters with health promotion programs. Although decisions to use programs in schools are typically made at the administrative level, teachers are the primary agents of school-based prevention efforts. Their support, motivation, and commitment are crucial to implementation success [1]. In the Netherlands, teachers are the ones who decide to use a school-based program in their classroom (adoption), deliver the program to students (implementation), and continue to do so in the long run (continuation) [2]. Many evidence-based programs consider the implementers of the programs, such as teachers, as "core" to the success of the program [3].

Implementation of school-based health promotion programs is, however, not optimal [4,5,6]. These programs are not implemented with sufficient strength and fidelity to produce measurable outcomes [7]. A monitoring of school-based interventions in the Netherlands showed that only "5%-10%" of teachers who have bought a program, implement it fully in accordance with the ideas of the program designer, resulting in reduced program effectiveness [8]. The behavior of program implementers is often an aspect that is overlooked or insufficiently considered in program development. There is a need for greater attention for quality of implementation [9].

Implementation is a process consisting of several phases, namely adoption, implementation, and continuation [10]. Teachers need support in every phase of the implementation process to enable them to effectively carry out the program in their lessons [11,12]. Most interventions were aimed at supporting teachers in the awareness and adoption of the program but little is known or created to support teachers in the implementation phase [4,13,9,14]. Support in this phase is crucial, however, for optimal program effectiveness [13].

Especially when it comes to providing school-based sex education, delivering such lessons is not a simple or obvious task; teachers, who are key to the success of such programs, not only require knowledge and a positive attitude but also certain skills and competencies to deliver a range of sensitive topics in these lessons and to deal with the difficulties encountered during implementation of the program. To prepare teachers for program use, specialized and effective training is necessary [15]. Although training often equips teachers with skills for correct implementation, it is not enough [16]. It remains important to provide teachers with more personal assistance and ongoing support and consultation during program delivery to ensure the quality of implementation [6,17,18,14]. This support needs to be of sufficient duration to achieve depth in teachers' skills and behavioral change throughout program delivery [19]. Paulussen et al [2] highlighted the importance of providing support before and during the implementation of a curriculum by way of training and technical and didactic assistance to ensure enduring success.

### The "Long Live Love" Program

In the Netherlands, Long Live Love (LLL) is the most widely used, effective school-based sexual education program, partly due to a successful dissemination strategy [20]. An earlier study on the implementation of LLL revealed that trainings from an external party, the Municipal Health Services (MHS), resulted in improved implementation of LLL by teachers [21]. Due to economic cutbacks, the supportive role of the Dutch MHS has recently been limited to predominantly stimulating dissemination and adoption of LLL and preparing teachers for initial implementation. They lack the capacity to provide intensive and long-term support [6]. In addition, MHS professionals lack the didactic expertise and skills to be appropriate role models for teachers in teaching skills for adequate implementation [22]. Teachers therefore need another form of support during implementation to compensate for the limitations of the MHS and to complement the existing dissemination strategy of LLL.

To contribute to the limited documentation of implementation interventions, this paper presents the systematic development of a Web-based coaching intervention, Lesgevenindeliefde.nl (teaching love). The website is part of a broader dissemination strategy and supports teachers in implementation of the school-based sexual education program, LLL. The Web-based coaching intervention aims at an optimal implementation, with completeness and fidelity, of LLL by teachers. As of date, no other Web-based coaching website to support teachers in delivering school-based sexual education is known in the Netherlands [23,24]. This paper will provide insight into teacher implementation of a school-based sex education program, LLL, and describe the complete cycle of development of this coaching website, from problem to solution. The website is developed applying intervention mapping (IM), a protocol to systematically develop interventions using theory and empirical evidence [25]. IM has proven to be an effective protocol in the development of various Web-based health promotion interventions [26,27,28].

## Methods

### Developing Effective Behavior Change Interventions

Intervention Mapping (IM) is a protocol for the development of theory- and evidence-based interventions. It maps the path from identification of a problem to the development of a solution. Although IM is presented as a series of 6 steps (see Figure 1), it is an iterative and cumulative process in which, respectively, the developer moves back and forth between the steps and in which each step is based on the outcomes of the previous ones [25]. The 6 steps are (1) conduct a needs assessment, (2) create matrices of change objectives, (3) select theory-based methods and practical applications, (4) organize methods and applications into an intervention program, (5) plan for adoption, implementation, and sustainability of the program, and (6) generate an evaluation plan [25].



XSL•FO
RenderX

### Intervention Mapping Steps

The first step, the needs assessment, begins with establishing a participatory planning group, represented by potential program participants and implementers. This step consists of a full analysis and description of the problem through an epidemiologic, behavioral, and social analysis of the at-risk-group. By means of qualitative and/or quantitative research, behaviors and environmental factors related to the health problem are identified.

In step 2, a transition is made from the problem to the solution, namely specifying what should change to prevent or to minimize a problem. Step 2 begins with the formulation of the behavioral and environmental outcomes to be achieved by the intervention followed by a breakdown of these outcomes into specific sub-behaviors called performance objectives, stating what the target group needs to do to achieve these desired outcomes. Next, determinants are selected that are linked to these objectives. Finally, these determinants and performance objectives are connected in a matrix to create change objectives, which state the specific goals to be achieved as a result of the intervention.

In step 3, theoretical methods are selected that change the specified determinants and achieve the change objectives. A method is a general technique for influencing change in determinants. These methods are translated into practical applications while taking the parameters for use into consideration. These parameters provide conditions under which effectiveness of the application is ensured. The applications should fit within the context of the intervention and the target group.

In step 4, creative and effective program components and materials are developed based on the previous steps. The challenge is to cover all selected theoretical methods, practical applications, and change objectives. The end product of this step is a coherent program that remains true to the planning that has been accomplished in step 1, 2, and 3.

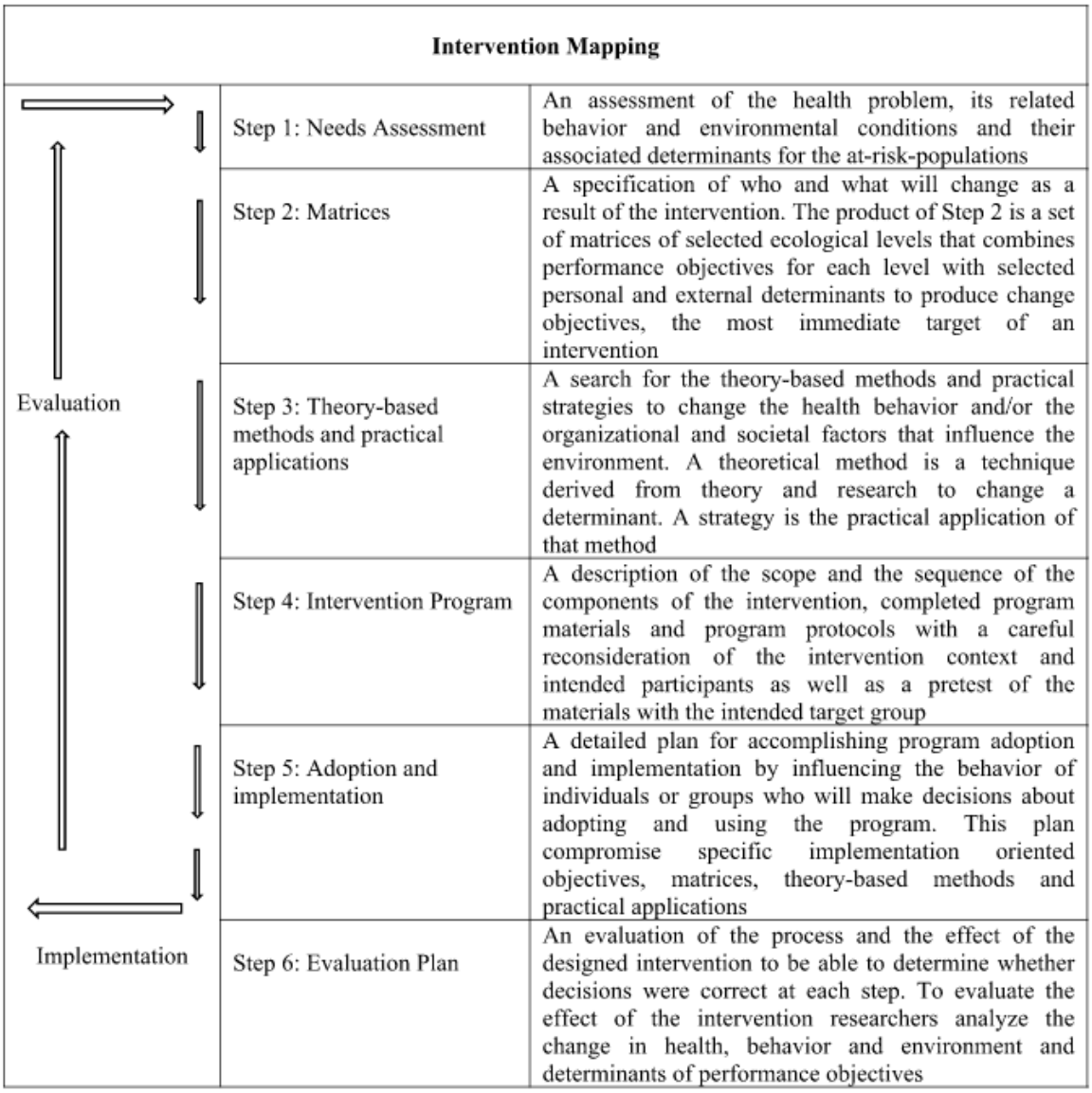
Effective programs, however, will have limited impact if they are never, incorrectly or incompletely used. An appropriate adoption, implementation, and maintenance plan is essential to achieve the desired outcomes. The main goal of step 5 is to ensure that the intervention will be used and maintained over time for as long as is needed. To realize this goal, the developer must complete the same tasks as in step 1, 2, 3, and 4, resulting in an effective intervention plan for optimal adoption, implementation, and continuation of the intervention.

In the final step of the iterative and cumulative IM process, the effect and implementation of the intervention are evaluated. An evaluation gives insight into the extent to which the earlier formulated goals are achieved after application of the intervention. The evaluation is divided in outcomes of quality of life, health, and behavior. A process evaluation is necessary to understand these outcomes and gives insight in the “black box” underlying the effect. The “black box” provides insight into what happens between application of the intervention and the outcomes. This paper presents outcomes of steps 1-5. The effect and process evaluation will be presented in a separate paper.



XSL•FO
RenderX

**Figure 1.** Intervention mapping process. Adapted from Bartholemew et al [25].




## Results

### IM Step 1: Needs Assessment

At the start of the project, a participatory planning group was set up, consisting of a panel of health promotion professionals (N=10), teachers (N=4), and MHS professionals (N=2). The goal for the selected group was to think along in the intervention development process and be consulted throughout the project. The needs assessment was conducted by means of (1) analyzing existing programs and reviewing the literature and (2) qualitative research.

#### *Analyzing Existing Programs and Literature Study*

The search for existing programs in the Netherlands did not reveal the existence of systematically developed and evidence-based Web-based interventions for coaching teachers in providing sexual reproductive health (SRH) lessons. The search did result in a few materials to support teachers in teaching SRH. This support was, however, minimal and not aimed at coaching to bring about behavioral change. In the field of sexual health promotion, for example, there is a website for teachers, but this is limited to providing materials and practical information on how to provide such lessons without further coaching [29]. This is insufficient for behavioral change, which is necessary for completeness and fidelity of program delivery [25,30,12].

The literature study revealed that limited examples are available on the development, execution, and evaluation of implementation enhancing interventions in general. In fields other than SRH, studies were also mainly focused on the provision of technical support [31,32,33,34,35,36,37,38]. These studies, however, were not aimed at coaching to bring about behavioral change. Although related to themes other than SRH, these studies reconfirm the limited existence of evidence-based coaching interventions and emphasize the importance of systematically developing an intervention to accomplish sustainable behavioral change. Supporting teachers during implementation will enable them to deliver the lessons as complete as possible (completeness) and according to previously formulated program goals (fidelity) for optimal effectiveness [11,12].



XSL•FO
RenderX

### *Qualitative Research*

Qualitative data were collected by conducting semistructured interviews with teachers to provide more insight into their implementation behavior and to get insight in their (perceived) need and preferences for coaching. A sample of 15 teachers from 12 different schools was selected from the database of schools who had bought the previous LLL program. The selection was made based on regional representation and gender. Teachers were asked to participate in the research by email. Furthermore, during the interview process, the snowball effect resulted in the involvement of 3 additional teacher respondents. The main reason for nonresponse was a lack of time. In the end, N=11 teachers signed up to participate for the interviews.

A topic list guided the interviews with 11 teachers (5 male, 6 female) from 9 different schools and regions, with diverse levels of experience in teaching SRH. The average duration of the interviews was approximately 40 minutes. See Textbox 1 for the topic list. This topic list was derived from a conceptual model based on the Theory of Planned Behavior [39], the Social Cognitive Theory [40], and the Diffusion of Innovations Theory [10]. These theories are often used to explain implementation behavior of teachers [21,2].

**Textbox 1.** Topic list needs assessment: Lesgevenindeliefde.nl

- What do you do with the theme of SRH?
- How do you teach your SRH lessons?
- Do you have any idea how other colleagues are dealing with the theme of SRH?
- Do you have any idea how less experienced colleagues are teaching SRH?
- Which difficulties do you experience in teaching SRH?
- How do you deal with those difficulties?
- How can you address those difficulties?
- What do you need to be able to teach SRH optimally?
- What do you need to be able to effectively teach Long Live Love (LLL)?
- How do you prepare your SRH–lessons or for teaching the Long Live Love program?
- Do you use any kind of support or a program during the application of the LLL program or the SRH lessons?
- What do you do in the evaluation of the SRH-lessons or the Long Live Love program?
- Which support would you like to receive in teaching SRH?
- Which support would other, and maybe less experienced, colleagues like to receive in teaching SRH?
- How should this support or coaching look like?
- How can this be implemented in an internet based coaching program?

The interviews revealed that teacher implementation of SRH programs, including LLL, is not optimal; various components of the program are selected and delivered, rather than completing the entire program and implementing it as intended by the program developers. Teachers describe their implementation behavior as making a selection of program components, adjusting the program with their own additions, not delivering the program in its entirety, limited use of the teacher manual, and a lack of planning, preparation, and evaluation. This suboptimal implementation behavior may lead to reduced program effectiveness [4,5].

> *T: "When we teach Long Live Love, we sometimes make our own additions and modifications. The program lacks practical assignments. It's mainly about reading and answering questions."*
>
> *I: "What is required to provide SRH programs optimally?"*
>
> *T: "More practical materials. Actually, teaching SRH is mainly reading and answering questions. We do improvise with other materials because the SRH program alone does not contain sufficient practical assignments."*

Teachers do not acknowledge their behavior as problematic; they do not see the importance of delivering the lessons with completeness and fidelity for achieving program effectiveness and therefore expressed minimal need for coaching. Although the perceived need for support in implementation was low, teachers did recognize several difficulties that may be encountered, especially, according to them, by less experienced colleagues, during the provision of SRH lessons.

> *T: "Some colleagues, not myself of course, experience difficulties in talking about sexuality. How do you begin? How are you going to talk about it or cope with it? Coaching could be given for those kinds of problems."*

An inability to adequately deal with these difficulties can negatively interfere with optimal implementation of SRH programs. According to the respondents, teachers providing SRH lessons may encounter the following difficulties: (1) barriers to creating a safe and trusted atmosphere in the



XSL•FO
RenderX

classroom; (2) feelings of shame or a closed attitude toward sexuality; (3) dealing with personal questions asked to them or to other students; (4) coping with individual student problems related to SRH; (5) dealing with homonegative reactions and behavior of students in the classroom; (6) anticipating on negative events on social media and Internet among students in relation to SRH; and (7) providing SRH lessons in a culturally, gender-wise, and sexually diverse classroom.

> *I: "What skills, knowledge, or other factors do you need to educate the students about SRH?"*
>
> *T: "You have to perceive the world as students do. You shouldn't be surprised by comments in the classroom. You shouldn't assume that they are not sexually active. But you also need skills to create a safe and secure atmosphere for the students. They also have to be able to talk freely about their experiences. That you can use these experiences to give information and to integrate this in the lessons. It is also really important that there is respect for each other and for each other's boundaries. That you are able to establish your boundaries. That you can be different. That's also an important focus of our lessons."*

When teachers were asked what was necessary for effective implementation of SRH programs, teachers mentioned a desire for materials they could use in the classroom with their students. They were not focused on their own quality of implementation but instead they were on the lookout for practical tools to use during lessons.

> *T: "Teachers don't often place their problems on a forum after a lesson. They will use it to find ideas for their lessons. Then they search tediously. But they won't share the experiences they had during a lesson on a forum."*
>
> *I: "How can we optimally coach a teacher so that he is capable of teaching Long Live Love or SRH?"*
>
> *T: "For teachers, it is necessary to be well informed about the topic. To have adequate, sufficient, and reliable information available. That they have the feeling; "I can answer questions." That's important in my opinion."*
>
> *I: "What more do teachers need?"*
>
> *T: "Materials and good information. Ehm… Something to visualize. The classical cucumber with a condom."*

When the teachers were directly asked about their need for coaching, the respondents answered to be satisfied with their own teaching method and expressed minimal need to be coached. They felt they could prepare sufficiently by reviewing the program materials and the teacher manual individually, or were incidentally assisted by other teachers in preparations for program implementation.

> *T: "Well, in my case, I don't know if I would use it (an e-coaching website) extensively. Because I have been teaching this (SRH) for a long time, I know a lot and everything can be talked about. If I don't know something, I go to a colleague. So, I don't know if I would make use of it. I would take a look. Purely out of curiosity. Maybe I am too arrogant but I really can't think of anything I would need help for."*

Critical reflection of one's own behavior and working on professional development are necessary for creating awareness and establishing sustainable behavioral change [41]. Teachers do not seem to see the connection between completeness, fidelity, and program effectiveness, do not see their suboptimal implementation behavior as problematic, or are not aware of potentially challenging situations and therefore do not see a need for behavioral change and coaching. Teachers need to be aware of the importance of completeness and fidelity in relation to program effectiveness, have insight in their (suboptimal) implementation behavior, and be aware of potentially challenging situations, to ultimately improve completeness and fidelity of program delivery. Critical self-reflection leads to awareness of own behavior and is the first step of coaching teachers toward behavioral change and professional development. Without a genuine recognition of need and desire, it is almost impossible to change behavior [41].

> *I: "Do you think teachers would make use of such a coaching website?"*
>
> *T: "If you point out the things that can go wrong, they have to prepare to deal with them. If you can trigger that, you've already come a long way. Teachers will start to reflect; "how does that affect me?""*

If a coaching intervention was to be developed, it is important, according to the teachers, to develop an intervention that is easy to use and accessible and does not cost a lot of time and effort because teachers claimed that they only have limited time and resources to prepare or to evaluate the lessons.

> *I: "How should such an e-coaching website look like?"*
>
> *T: "There shouldn't be any complicated access codes. A lot of people drop out if they see that. It should be easily accessible. It shouldn't cost me an hour and a half to browse. I don't have time for that. Ideally you can select various components on a website while browsing; difficult situations that you may encounter. If a recognizable situation is described by a fellow teacher, I might think, this can happen to me as well."*

The possibility of developing virtual coaching, in which teachers are guided throughout the implementation process by a virtual buddy, was discouraged by most of the respondents. Instead of a virtual buddy, teachers expressed the preference to communicate with colleagues within different schools to exchange ideas and teaching methods or to solve problems they encounter during the provision of SRH lessons.

> *I: "Do you evaluate or discuss your lessons?"*
>
> *T: "No, it is a very lonely profession… It is progressive thinking to learn from other teachers."*

In conclusion, the needs assessment revealed an interesting finding: there is a discrepancy between teachers perceived and actual need for support in providing SRH lessons effectively. Teachers do not perceive their implementation behavior as



XSL•FO
RenderX

problematic, but their actual implementation behavior does not fulfill the required completeness and fidelity for program effectiveness. To ensure fidelity and completeness of program implementation, it remains important to first create awareness, by means of self-reflection, of (1) the importance of completeness and fidelity in program implementation, (2) teachers' current implementation behavior, and (3) the difficult situations they could potentially encounter. To achieve behavioral change, and contribute to professional development, teachers should subsequently be supported in dealing with the common difficulties mentioned and be provided with the knowledge and skills they need to implement SRH programs effectively. A careful choice for unobtrusive coaching techniques was made to ultimately bridge the gap between the perceived need and actual need of teachers for coaching. The technique of unobtrusive coaching is required to create awareness and accomplish behavioral change, despite the teacher's resistance to coaching and ultimately optimize the role of the teacher in providing high-quality SRH lessons.

### IM Step 2: Matrices of Change Objectives

Based on the needs assessment and literature review, a program goal was formulated and subdivided into 4 desired behavioral outcomes for teachers. The program goal was that teachers in all secondary schools in the Netherlands implement LLL completely and according to its formulated goals (fidelity). The behaviors associated with this program goal were that teachers (1) reflect critically on and become aware of their implementation behavior regarding SRH, (2) implement LLL completely, (3) implement LLL according to guidelines in the teacher manual, and (4) deal adequately with difficulties that may be encountered during provision of SRH. These behaviors formed the outcomes of the intervention and were subsequently broken down into performance objectives. Performance objectives specified what teachers needed to do to perform those desired behaviors. The formulated behavioral outcomes and related performance objectives are presented in Table 1.

Performance objectives were then linked with their associated personal and external determinants. Determinants were specified based on the results of the needs assessment, a literature review and a review of theories. Social influence was not selected as a determinant because the interviews revealed that teachers individually determined their own method of teaching. However, skills, self-efficacy, attitude, and knowledge were found to be important determinants for teachers' implementation behavior [6]. These determinants were evaluated on importance (how strongly is the determinant related to teacher's performance objectives) and changeability (how easily can the determinant be influenced by a theory-based method), which formed the basis for the final selection of determinants that the intervention would target. A matrix was then created by combining the performance objectives and associated determinants to create change objectives; specific and measurable goals specifying what will change among teachers as a result of the intervention. For example, "teachers express confidence (determinant: self-efficacy) in creating a safe and secure atmosphere in the classroom when delivering LLL (performance objective)". See Table 2 for a selection of change objectives.

### IM Step 3: Theory-Based Methods and Practical Applications

In this step, we selected theory-based methods to change the specified determinants and ultimately achieve the change objectives. These methods were derived from theories, predominantly the Social Cognitive Theory, Elaboration Likelihood Model, and the Trans Theoretical Model and from evidence in the empirical literature stating that the methods might have the desired effect to change the determinant [25]. The parameters, the conditions under which the methods were expected to be effective, were considered when translating them into practical applications, which fit within the context of the intervention and target group. Table 3 shows examples of selected theoretical methods, practical applications, and their relation to the selected determinants. For example, behavioral journalism is a potentially effective method for increasing self-efficacy but will only work under the condition that authentic interviews are used with actual community members, which represent the desired message [42]. This method was translated into the application of role model stories where teachers share their experiences and suggestions in dealing with difficult situations.



XSL•FO
RenderX

**Table 1.** Behavioral outcomes and performance objectives of teacher implementation.

| Behavioral outcomes of teachers | Performance objectives |
|---|---|
| B.O.1. Teachers reflect and improve on their implementation behavior regarding SRH | P.O.1. Teachers reflect critically on their implementation behavior regarding SRH |
| | P.O.2. Teachers self-monitor and improve the weaknesses in their implementation behavior regarding SRH |
| B.O.2. Teachers deliver LLL to students completely (completeness=80% of program) | P.O.2.1. Teachers cover all 6 lessons of LLL |
| | P.O.2.2. Teachers use all program materials of LLL in each lesson (DVD, magazine, teacher manual, worksheets) |
| | P.O.2.3. Teacher covers the most important components of each lesson, as indicated in the teacher manual |
| B.O.3. Teachers deliver LLL to students according to the guidelines in the teacher manual (fidelity) | P.O.3.1. Teachers read the teacher manual as preparation for each lesson |
| | P.O.3.2. Teachers deliver each LLL lesson to students according to the teacher manual |
| B.O.4. Teachers deal adequately with the most common difficulties that arise during implementation of SRH | P.O.4.1. Teachers create a safe and trusted atmosphere in the classroom during all LLL lessons whereby students feel comfortable in the classroom and asking questions |
| | P.O.4.2. Teachers teach all themes in LLL without shame or taboos interfering with the quality of the lesson |
| | P.O.4.3. Teachers handle personal questions of students addressed to themselves depending on their personal need to answer these questions |
| | P.O.4.3.1 Teachers intervene whenever students ask them or fellow students questions that are too personal |
| | P.O.4.4. Teachers integrate the theme of homosexuality as self-evident during all lessons of LLL |
| | P.O.4.4.1 Teachers intervene on homonegative behavior of students |
| | P.O.4.5. Teacher handle cultural, gender, and sexual experience diversity in the classroom using an approach that address and involves all students |
| | P.O.4.6. Teachers identify individual problems of students with and refer them to the appropriate help |
| | P.O.4.7. Teachers address actual themes within social media and Internet in relation to SRH during the provision of LLL |
| | P.O.4.8. Teachers facilitate discussions in the classroom about relationships and sexuality according to the formulated goals and suggestions in the teacher manual |



XSL•FO
RenderX

**Table 2.** Sample of change objectives

| Homosexuality | Knowledge | Attitude | Self-efficacy | Skills |
|---|---|---|---|---|
| 1. The teacher integrates the theme of homosexuality as self-evident during all lessons of Long Live Love | K 1.1 The teacher describes how homosexuality is integrated in the lessons of Long Live Love. | A 1.1 The teacher expresses the importance of a positive attitude of a teacher toward homosexuality during the application of the lessons of Long Live Love. | SE 1.1 The teacher expresses confidence in ability to replace "he" and "she" by "he" and "he" or "she" and "she." | S 1.1 The teacher demonstrates how he/she continually integrates the theme of homosexuality in the lessons. |
| | K 1.2 The teacher lists the moments in the Long Live Love lessons where the theme of homosexuality can be discussed as a self-evident part of the lesson. | A 1.2 The teacher expresses the advantages of integrating homosexuality as self-evident during the application of Long Live Love. | SE 1.2 The teacher expresses confidence in the ability to continually integrate homosexuality and certainly not avoid the theme in the lessons of Long Live Love in case of negative reactions from students. | S 1.2 The teacher demonstrates skill to not avoid the theme of homosexuality despite possible adverse or negative reactions from students. |
| | K 1.3 The teacher describes how he/she plans to integrate homosexuality in the lessons of Long Live Love. | A 1.3 The teacher expresses the importance of mentioning "he" and "he" or "she" and "she" instead of "he" and "she" during the lessons of Long Live Love. | SE 1.3 The teacher expresses confidence in the ability to protect students with feelings of homosexuality against a feeling of discomfort or social pressure. | S 1.3 The teacher demonstrates skills to stimulate the discussions about homosexuality in which respect and acceptance are important key aspects in this in depth discussion. |
| | K 1.4 The teacher explains that when "he" and "she" is mentioned this can also be replaced by "he" and "he" or "she" and "she." | A 1.4 The teacher expresses the importance of discussing and integrating the theme of homosexuality, especially when the students react negatively. | SE 1.4 The teacher expresses confidence in the ability to communicate the message that homosexuality is not limited to borders, cultures, or countries during the lessons. | S 1.4 The teacher demonstrates how he/she protects students with homosexual feelings from a feeling of discomfort. |
| | K 1.5 The teacher explains the reasons why he or she will strive toward a self-evident integration of homosexuality as theme in the lessons of Long Live Love. | A 1.5 The teacher describes the importance of effectively coping with feelings of pressure or discomfort of students with feelings of homosexuality during the lessons of Long Live Love. | SE 1.5 The teacher expresses confidence in the ability to be continually alert of possible individual confrontations between students about homosexuality. | |
| | | A 1.6 The teacher expresses the importance of informing students that homosexuality is not limited to a culture, to borders, or to periods. | | |
| 2. Teachers intervene on homonegative behavior of students | K 2.1 The teacher lists the signs he/she should be aware of which could indicate homonegative ideas or behavior among students. | A 2.1 The teacher expresses a disapproving attitude toward homonegative behavior during the application of Long Live Love. | SE 2.1 The teacher expresses confidence in the ability to be continually alert of signs or behavior of students in the classroom, which can be homonegative. | S 2.1 The teacher demonstrates skills to constantly be alert of homonegative signs or behavior of students during the lessons. |
| | K 2.2 The teacher describes which methods can be used effectively in the classroom when students have homonegative ideas or show homonegative behaviors. | A 2.2 The teacher describes the importance of being constantly alert of homonegative signs or behavior of the students. | SE 2.2 The teacher expresses confidence in ability to take measures when students act homonegatively in the classroom. | S 2.2 The teacher demonstrates skills to adequately deal with homonegative signs or behavior of students in the classroom. |
| | K 2.3 The teacher describes how homonegative reactions of students can be used as a subject for discussion. | A 2.3 The teacher expresses the importance of taking timely measures when students act homonegatively in the classroom. | | |



XSL•FO
RenderX

**Table 3.** Methods and applications.[a]

| Determinants | Methods | Parameters | Applications | How population, context, and parameters were taken into account |
|---|---|---|---|---|
| Knowledge | Elaboration | High motivation and cognitive ability, personally relevant messages, surprising, repeated, self-pacing, not distracting, easily understandable, include direct instructions | Informative texts, tips, and FAQ | *Population:* The informative text was derived from professional teacher channels as well as from teachers themselves and health promotion, didactic, and pedagogic professional information.<br>*Context:* Texts and tips were included within each sub homepage for each difficulty.<br>*Parameters:* The texts were revised by an editor, composed based on teachers' experiences and relevant literature for the area of expertise. |
| | Feedback | Specific, follows behavior in time, individual | Email option and options to post reactions on role-model stories and films | *Population:* To be able to answer specific individual questions and provide individual feedback this function was integrated in the website.<br>*Context:* In case of a direct coaching question from the visiting teacher, he/she was able to ask questions through a mail-function or post comments below a story or film.<br>*Parameters:* The mail form was only accessible for the individual teacher. The question or answer was not published for others. Continuation in contact could be initiated by the teacher. |
| | Discussion | Listening to learner to ensure correct schemas are activated | Options to post comments on role-model stories and films | *Population:* Visiting teachers were stimulated by an open question to post comments. The open question structured the discussion.<br>*Context:* The aim of the comments below the role-model stories and the film was to stimulate a discussion between visiting teachers about the topic discussed in the story or film. It also served as a platform for discussing tips on how to deal with that specific difficulty.<br>*Parameters:* The placed reactions were visible for all visiting teachers. A Youtube-like structure was used for optimal usability. |
| Skills/self-efficacy | Behavioral journalism | Credible message, model gives reasons for adopting new behavior and states perceived reinforcing outcomes received | Rotating photo's, role-model stories and films | *Population:* Interviews with teachers were used in several aspects of the website to realize a platform by and for teachers.<br>*Context:* Photo's and interviews were used to compose role-model stories, films, and photos. These stories and films were based on a structure in which first the problem is presented as well as the experience and the relevance of this problem followed by the search for the most effective solution with a description of failures and success factors.<br>*Parameters:* The interviewed teachers were selected to present a diverse selection in teaching experience, in geographic location and personal characteristics and were coping models, instead of mastery models, to increase the identification. |
| | Modeling | Attention, remembrance, self-efficacy and skills, reinforcement of the model, identification with model, coping instead of mastery model, demonstrate relevant skills | Rotating photos, role-model stories, and films | *Population*:To create a platform for and by teachers, teachers were interviewed which formed the content for role-model stories and films. Photos of teachers were taken to increase reliability and credibility as well as to lure teachers to the website.<br>*Context*:The interviews were used to fill in the main content of the website.<br>*Parameters*:Interviewed teachers were selected on personal characteristics, on geographic location, and on experience to create a database of diverse teachers that the target group could identify with. The interviewed teachers were all coping models. |



XSL•FO
RenderX

| Determinants | Methods | Parameters | Applications | How population, context, and parameters were taken into account |
|---|---|---|---|---|
| Attitude | Self-reevaluation | Feedback and confrontation; however, raising awareness must be followed by increase in problem solving ability and self-efficacy | Self-reflection tool | *Population*: Teachers expressed a minimal need for coaching but teachers' implementation behavior was characterized by inconsistently selecting parts of the program and not delivering all lessons. To bring about behavioral change first a self-reflection intervention is necessary to create professional awareness as the first step of improving implementation behavior.<br>*Context*:In this self-reflection tool teachers could score different aspects of their own professional behavior in dealing with difficulties on a Likert scale.<br>*Parameters:* After completion of the self-reflection tool an overview of gaps in learning were revealed. The teachers were directly referred to the most personal relevant difficulties. |
| | Scenario-based information | Plausible scenario with cause and outcome, imagery | Role-model stories and films | *Population:* Teachers were interviewed to collect data and to form the content for the scenarios. Teachers were coping models who were also experiencing the same problems as the target group but also found a solution.<br>*Context:* In the films and stories, interviewed teachers were especially asked to describe scenarios to make the learning process applicable in daily practice and for the individual situation of a teacher.<br>*Parameters*: The scenarios were described according to a structure in which the (personal) relevance and description of the problem was made followed by a search for the most effective solution. |
| | Modeling (see above *)* | | | |

[a]Bartholemew et al [25]

## IM Step 4: Program Development

In this step, the intervention program is designed and materials are developed. The intervention, called Lesgevenindeliefde.nl (teaching love), was designed as a coaching website for teachers.

Although teachers expressed a minimal need for coaching during the interviews in the needs assessment, the program developers nonetheless saw the need to develop Lesgevenindeliefde.nl for the following reasons: (1) quality of implementation by teachers is suboptimal, despite their conviction about their own teaching method. To change this conviction and improve teacher implementation behavior, self-reflection and professional development are necessary. Teachers do not usually reflect on their own implementation behavior. Self-reflection, however, could lead to a critical evaluation and subsequently to improvement of their implementation behavior, which ultimately contributes to an increased effectiveness of an intervention. Coaching can only start when teachers develop an awareness of the need and a desire to improve their performance or change the way they have been performing at work [41]. Confronting teachers with potential difficulties they could encounter might help them to reflect on their behavior. (2) Teachers could use support in adequately dealing with the difficult situations encountered when delivering SRH lessons, as mentioned in the needs assessment. This could lead to improved program implementation. (3) The development of a website is an efficient, low-threshold way of reaching a mass of teachers. It partly replaces and supports the implementation promoting tasks of the MHS, which now lacks the capacity and didactic expertise for training teachers. The choice of the Internet as a channel of the intervention was predetermined by the program financers.

To ensure that the intervention was appealing and trustworthy to teachers, the coaching website was based on the concept "by and for teachers," with role model stories, photos, and videos as the main products of this concept. This concept was chosen because teachers stated that if they did seek support during the delivery of SRH lessons, they preferred to consult fellow teachers or considered other teachers as reliable sources of information and for seeking advice. A large study in the Netherlands found that teachers in secondary schools either consult colleagues in their school for information or use the Internet to find information, to prepare their lessons, send emails to students, or give homework assignments [23]. Studies in the United Kingdom pointed clearly to the value of teachers learning with and from each other when it comes to professional development [43,44].

Certain characteristics of the website such as its accessibility, usability, flexibility, and tailorability to needs of teachers potentially limited barriers to visiting the website. Teachers could flexibly access all the information on the website that was personally relevant rather than being forced through a fixed coaching program. This catered to their lack of time and diverse needs for support. Accessibility to the website was simplified by placing the link on the LLL e-learning website under the "teacher" button. Teachers were encouraged to come back to



XSL•FO
RenderX

the website by constantly placing new updates there and by integrating teacher materials in the website.

The Web-based intervention was developed with the underlying idea of an unobtrusive coaching technique whereby teachers' actual needs were addressed and their perceived need of coaching was changed. With this technique, we attempted to trigger teachers to become aware of their need for coaching, without awakening resistance, by exposing them to difficulties experienced by other teachers they identified with. To meet teachers' need for student materials, teaching materials were provided via the website. This student material can function as the first trigger for teachers to visit the website but was strategically placed at the bottom of the navigation system to ensure that teachers were first exposed to the most common difficulties encountered during implementation. It is a necessity that teachers effectively cope with difficulties to prevent these from becoming a barrier to optimal delivery of the program. Subsequently, elimination of these barriers is followed by practical support in delivering the lessons to ultimately accomplish completeness and fidelity of the delivery of LLL.

Furthermore, teachers were lured into the deeper structure of the website by presenting clickable rotating quotes with photos of teachers they could identify with. This is in line with the concept "by and for" teachers derived for the method of peer coaching. Peer coaching suggests that the professional development of teachers can be improved through experimentation, observation, reflection, the exchange of professional ideas, and shared problem - solving [45]. The information on the website was given by teachers instead of experts as they are coping models, not mastery models, which is important for the acceptance of the message [25].

The homepage of the coaching website includes a left-menu structure and rotating photos of teachers with SRH teaching–related quotes. These photos of teachers with short rotating quotes, placed in the center of the homepage, were meant to increase teachers' awareness of the most common difficulties and to trigger their perceived need to be coached. These quotes also served as cues for teachers to browse further through the website and as an entrance to the related role model stories. Furthermore, on the homepage, general information about the website could be found as well as the possibility to ask for support via email, up-to date information about SRH, a search function, and frequently asked questions (see Figure 2). Additional subpages could be reached by clicking on any of the features presented on the homepage.

To access the website, teachers had to sign up with a self-created user name and password. The sign-up was included to protect the privacy of teachers and to provide a protected Internet environment where teachers could safely exchange their ideas. The website also needed to be secured to prevent students from accessing it. For the program developers, these account details provided demographic information about the users and the use of the website.

Access to the main content of the website was predominantly navigated through the left-menu structure. This structure contains three categories: (1) a self-reflection tool to trigger teachers to reflect critically on their implementation behavior, (2) the 8 main difficulties that teachers may face when implementing SRH lessons, each with their own underlying sub-homepage, and (3) student materials and practical instructions teachers need to implement LLL completely and with fidelity. See Figure 2 for an impression of the homepage.

The left-menu structure was chosen for usability purposes, mainly because it presented the website content and structure clearly. This made the website accessible and easy to use. Current practices strongly recommend placing the main navigation menu on the left-hand side of the page [46]. The choice of this navigation system addressed teacher's limitations of time and skills in the Internet use and the general preference of website visitors to be able to have a clear navigation on the homepage [47].

The first part of the left menu structure consisted of a self-reflection tool. This was developed to enable teachers to reflect critically on their own implementation behavior and to create awareness of their need for coaching concerning the most common difficulties in providing SRH. The self-reflection tool was developed in the form of an interactive questionnaire whereby each statement related to teaching SRH was rated on a Likert scale of 1 (totally disagree) to 7 (totally agree). For example on the difficulty of teaching SRH without shame, the following statement had to be rated from 1 to 7: "I can deal with feelings of shame in such a way that they do not limit my lesson" (see Figure 3). With the results of the reflection tool, teachers were referred to personally relevant pages on the website to enable them to improve these specific behaviors.

The central component of the left menu structure focused on the 8 main difficulties encountered during implementation by presenting them via 8 separate buttons, each with its own sub-homepage. The difficulties were (1) creating a safe and secure atmosphere for the provision of SRH lessons, (2) teaching SRH without shame influencing the quality of the lessons, (3) protecting boundaries in sharing personal information and questions between teachers and students and among students themselves, (4) identifying personal problems of students, (5) integrating social media in SRH lessons, and (6) integrating homosexuality as self-evident during the lessons of LLL and dealing with negative reactions to homosexuality and anticipating on (7) differences in culture, and (8) different levels of sexual experience in the classroom.

By clicking on one of these menu items, the visitor entered a sub-homepage with background information on the specific difficulty, videos, role model stories with rotating quotes, and suggestions to adequately deal with that specific difficulty. The rotating quotes served the same function as the quotes on the homepage, namely to make them aware of the personal relevance of the difficulty, awaken their need to adequately deal with that difficulty and trigger them to access further content. Each role model story and video had the possibility of posting a comment, similar to existing commentary structures, such as YouTube (See Figure 4).

The content of the website was mainly obtained by interviewing teachers for role model stories and videos. It was important that the video's and role-model stories were realistic; teachers had to be able to identify with them, the content had to be

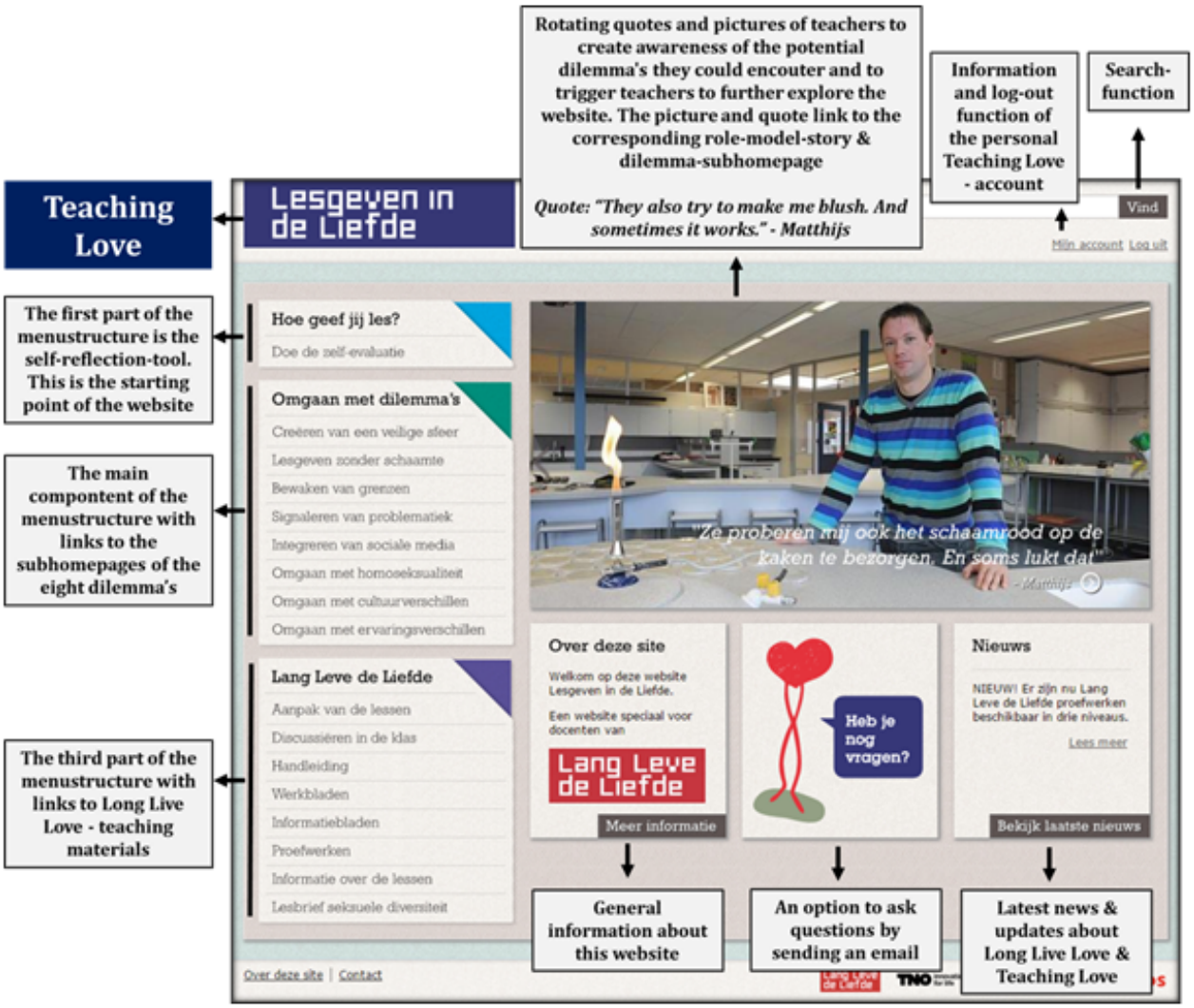


XSL•FO
RenderX

recognizable, and a clear solution on how to cope with the difficulty had to be provided. Each video and role model story was based on the structure of first introducing the teacher to the difficulty and then providing a solution. First knowledge, then awareness, attitude, and then skills were addressed in these videos and role model stories. The role model first introduced and acknowledged the difficulty (knowledge and awareness), then expressed the importance of dealing adequately with the difficulty (attitude) and finally described how (s)he effectively coped with the difficulty in such a way that it did not influence the implementation of LLL (see example of role model stories in Figure 5). The role model stories and videos were supplemented by general suggestions including didactic and pedagogic information and background information concerning the difficulty. This information was collected from websites and articles as well as from own research.

The last part of the left-menu structure was specifically aimed at effective practical delivery of the LLL program. Optimal delivery was achieved by informing teachers how to best provide LLL lessons and how to handle discussions in the classroom by means of role model stories and videos. Teachers could also download teaching materials such as the teacher manual, work, and information sheets for students and general information about the LLL curriculum. These materials were strategically included in the website to lure teachers to the website and meet the need they expressed in the needs assessment for such teaching tools.

**Figure 2.** Homepage of Lesgevenindeliefde.nl.






XSL•FO
RenderX

**Figure 3.** The self-reflection tool.



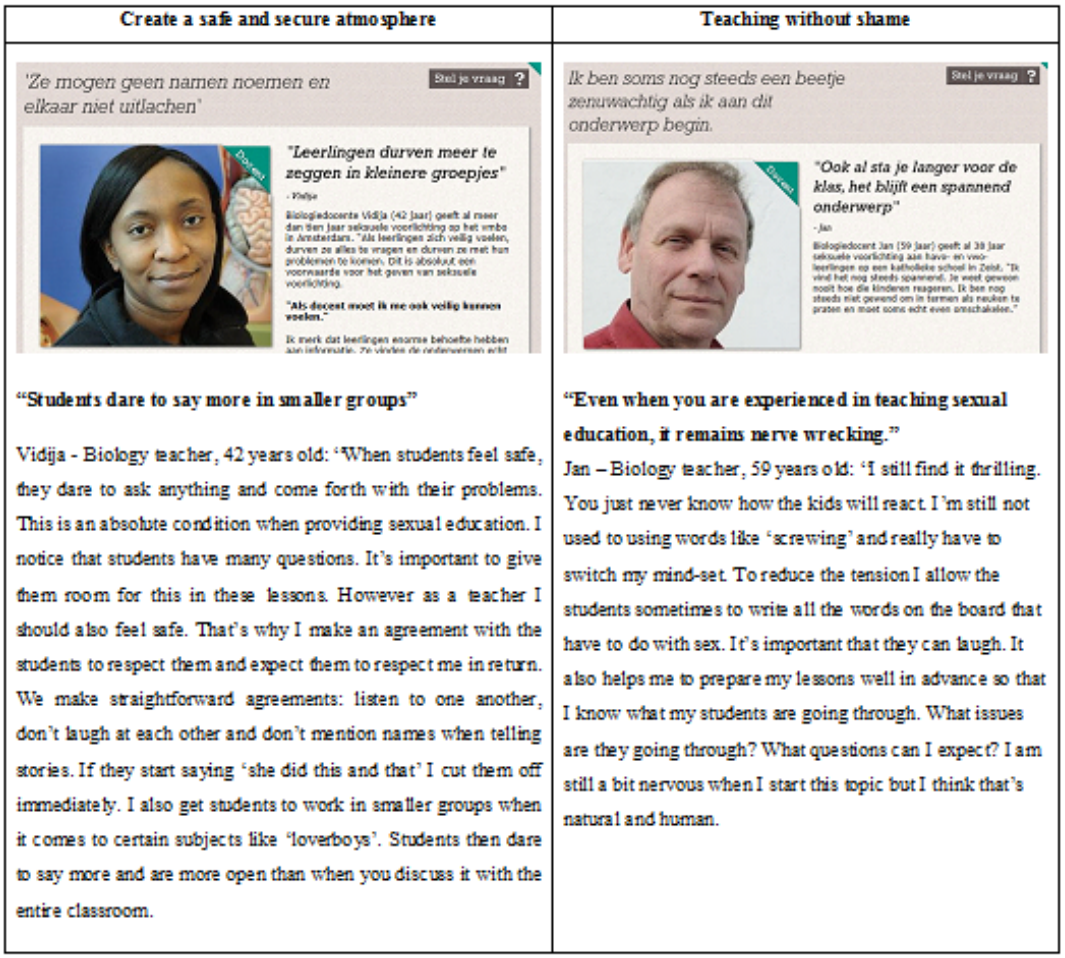

**Figure 4.** Subhomepage: Dealing with homosexuality.






XSL•FO
RenderX

**Figure 5.** Examples of role model stories.




## IM Step 5: Implementation

The coaching website (Lesgevenindeliefde.nl) had to be used by teachers to have an impact on program effectiveness and ultimately on student outcomes. In this step, an implementation plan was made to ensure that teachers were aware of the existence of the website and made use of it. Despite being designed to support teachers in their implementation of LLL, the website itself also needed to be effectively implemented.

The implementers of the coaching website are the program developers, STI AIDS Netherlands, who maintained, monitored, and updated the website and made it available and easily accessible on the Internet. Teachers were the end users of the website. The implementers developed dissemination tools, according to the IM protocol, to ensure that teachers were exposed to the website and to increase awareness and use of the website. An information brochure including information about LLL and the website was created. In addition, a trailer of the website was developed, explaining the aim of the website and showing the content and use of it (see Multimedia Appendix 1). This trailer and further information about the website were also integrated into the training provided to teachers by the MHS. In this training, teachers were informed by the MHS about LLL as well as the existence, advantages, and use of the coaching website, thereby stimulating teachers to use it when implementing LLL.

To further ensure teachers' awareness of the website and ability to make use of it in an efficient manner, a link to the website was integrated into the teacher manual. In the manual, references were made to the website in each lesson where relevant or wherever a specific difficulty was expected to arise in that lesson. The aim and functionalities of the website were also described in the teacher manual.

## IM Step 6: Evaluation

In this step of IM, an evaluation plan and the corresponding evaluation measures were identified and developed. An effect and process evaluation for the pilot implementation of the coaching website, Lesgevenindeliefde.nl, was performed. This occurred simultaneously with the pilot implementation of the school-based LLL intervention for students [20]. Formulated outcomes of steps 2 and 5 of IM, namely the change objectives and the implementation goals, were used in creating the evaluation plan. The aim is to find out how and to what extent teachers made use of the website, how they appreciated it, what effects it had on their implementation of LLL (completeness and fidelity), and what factors affected teachers use of the website. This was investigated using a randomized controlled trial design. Qualitative and quantitative data were collected. Results of this study will be described elsewhere.



XSL•FO
RenderX

## Discussion

### Principal Findings

In this paper, the Intervention Mapping (IM) protocol was applied for the development of a theory- and evidence-based Web-based coaching intervention, Lesgevenindeliefde.nl (teaching love), aimed at supporting teachers in their implementation of the sexual education program, Long Live Love (LLL). The IM protocol is perceived as a useful tool for guiding the development of this intervention and making it one which is compatible with the needs and preferences of teachers. With the IM protocol, careful decisions were made using a cumulative and iterative process, resulting in this Web-based implementation support intervention by and for teachers. Although IM was useful in designing this intervention, it is also a time-intensive method, which was exacerbated by limited available resources in the project. Predetermined requirements of the program financers, such as budget and time, and available capacity and time of the MHS restricted the options for the type of intervention such as a digital versus a personal form of coaching. The method of IM used to develop the e-coaching intervention can be applied in other school settings or extrapolated to other areas of health promotion [25].

The first step in the IM process revealed an interesting and challenging discrepancy between teachers' actual implementation behavior and their perception of their behavior. Several studies revealed that there is an implementation problem among teachers but teachers themselves do not perceive this suboptimal implementation behavior as a problem [4,5,48,16,49]. In the case of LLL, teachers did not deliver lessons completely and according to the goals of program designers but teachers did not perceive this behavior as problematic or as a necessity to effectively provide the lessons. This behavior, however, could possibly result in decreased program effectiveness. Teachers did, however, acknowledge some difficulties encountered when providing sexual education, who according to more experienced teachers, were predominantly faced by less experienced colleagues. These difficulties could affect the fidelity and completeness if they are not adequately dealt with.

Teachers remain a difficult target group for bringing about behavioral change and made it challenging to develop a coaching intervention. Although the intervention was necessary, teachers expressed resistance and reluctance and experienced a limited intrinsic motivation to be coached, thereby rendering directive coaching as a noneffective method for behavioral change of this target group. Directive coaching has been found to be an effective method to expose teachers to the intervention content but eagerness is a necessity before entering and being involved in this directive coaching process [47]. Lack of intrinsic motivation namely forms a barrier to coaching [50]. Without a need and desire to be coached, it is almost impossible to change behavior. Coaching can only start when teachers develop an awareness of the need and desire to improve their performance or change the way they have been doing things at work [41]. This resulted in selecting the technique of unobtrusive coaching to create awareness without creating resistance; a nondirective way to bring about behavioral change. To achieve behavioral change, self-reflection is an essential first step to create intrinsic motivation to be coached and to ultimately develop professionalism in SRH. In addition, a concept of peer coaching was integrated, by and for teachers, to lower the resistance and to ultimately achieve behavioral change. Peer coaching suggests that the professional development of teachers can be improved through experimentation, observation, reflection, the exchange of professional ideas, and shared problem solving [45].

To lure teachers to the website, student materials were made available on the Web. The lack of intrinsic motivation to visit the website triggered program developers to invest in additional implementation activities because a website alone would be insufficient to involve teachers who are not intrinsically motivated. Integrating the website in the teacher manual, incorporating information about the website in teacher training, and developing a trailer to create awareness and enthusiasm among teachers for the website are examples of implementation activities.

The increased use of digital technologies in the education system, such as Lesgevenindeliefde.nl, brings exciting opportunities for innovative ways of teaching and learning. New, Web-based technologies do not only provide an anonymous communication space but also offer students and teachers easier, affordable, convenient, and faster access to information, teaching and learning resources, peers, experts, and a wider community. Web-based technology is also a low-threshold and efficient way of reaching many teachers and providing support in, for example, the implementation of school-based programs. Exploring the educational potential of these digital technologies and supporting schools in making optimum use of them remains important [51].

### Conclusions

With the development of the e-coaching website, a unique contribution was made in the field of bringing about behavioral change among intermediaries, especially due to the elements of self-reflection and unobtrusive peer coaching. The use of Web-based coaching to improve implementation behavior of teachers could be generalized to different cultural contexts because it addresses the common challenges faced in the area of sexual health education in schools worldwide [52]. Our process of intervention development may be applied to get from problem to solution regarding diverse implementation problems in development of interventions for challenging target groups. In addition, other health promoting professionals may benefit from our example of the ongoing process of balancing input of the target group with the wishes of the intervention developers to ultimately develop an effective intervention.

Lesgevenindeliefde.nl will be tried out in practice by means of a pilot implementation. During this pilot-implementation, the website will be evaluated on process and effect. Based on the experiences of teachers and outcomes of the studies, further enhancements of the website could be made. The introduction of an innovation, such as Lesgevenindeliefde.nl, could present certain challenges in the implementation phase. The process of accepting the innovation takes time, as described in Rogers' diffusion curve [53]. The innovation will most likely be adopted by innovators first, followed by the early majority and eventually



XSL•FO
RenderX

the laggards. Pijpers et al [54] claim that Web-based innovations first need to be accepted to be used broadly and effectively. The systematic approach and customized concept of Lesgevenideliefde.nl can serve as a distinct example of how to bring about behavioral change in a target group of intermediaries who lack intrinsic motivation for and have resistance to coaching. This approach can be applied to other SRH programs in the school setting and to target intermediaries in health promotion.

## Acknowledgments

This work was funded by ZonMw and supported by Maastricht University and STI Aids, The Netherlands. The authors thank all the teachers who participated in the research.

## Conflicts of Interest

None declared.

## Multimedia Appendix 1

Trailer of 'Lesgevenindeliefde.nl'.

[MOV File, 241MB - jmir_v18i7e136_app1.mov ]

## References

1. Hunter L, Elias MJ, Norris J. School-Based Violence Prevention. Journal of School Psychology 2001 Mar;39(2):161-175. [doi: 10.1016/S0022-4405(01)00058-9]
2. Paulussen T, Kok G, Schaalma H, Parcel GS. Diffusion of AIDS curricula among Dutch secondary school teachers. Health Educ Q 1995 May;22(2):227-243. [Medline: 7622390]
3. Coyle K, Firpo-Tiplett R, Taylor J, Rex P. Scotts Valley, CATR Associates. 2011. Draw the line/respect the line Adaptation Kit URL: http://recapp.etr.org/recapp/documents/programs/DTL_Adaptation_Kit.pdf [accessed 2015-08-18] [WebCite Cache ID 6aryxbPwS]
4. Bessems KM, van Assema P, de Vries NK, Paulussen TW. Exploring determinants of completeness of implementation and continuation of a Dutch school-based healthy diet promotion programme. International Journal of Health Promotion and Education 2014 May 23;52(6):315-327. [doi: 10.1080/14635240.2014.912445]
5. Peters LWH, Kok G, Ten Dam GTM, Buijs GJ, Paulussen TW. Effective elements of school health promotion across behavioral domains: a systematic review of reviews. BMC Public Health 2009;9:182 [FREE Full text] [doi: 10.1186/1471-2458-9-182] [Medline: 19523195]
6. Schutte L, Meertens RM, Mevissen Fraukje EF, Schaalma H, Meijer S, Kok G. Long Live Love. The implementation of a school-based sex-education program in The Netherlands. Health Educ Res 2014 Aug;29(4):583-597 [FREE Full text] [doi: 10.1093/her/cyu021] [Medline: 24817522]
7. Gottfredson DC, Gottfredson GD. Quality of School-Based Prevention Programs: Results from a National Survey. Journal of Research in Crime and Delinquency 2002 Feb 01;39(1):3-35. [doi: 10.1177/002242780203900101]
8. Hoekstra F, Velderman MK, Wiefferink K, van Overveld C, ter Beek J, Buijs G, et al. De adoptiefase van schoolgerichte preventieprogrammas. Van boodschap naar beslissing. : Tijdschrift voor Gezondheidswetenschappen (TSG); 2009:27-36.
9. Fagan AA, Mihalic S. Strategies for enhancing the adoption of school-based prevention programs: Lessons learned from the Blueprints for Violence Prevention replications of the Life Skills Training program. J Community Psychol 2003;31(3):235-253. [doi: 10.1002/jcop.10045]
10. Rogers E. Diffusion of innovations. 3rd edition. New York: Free Press; 1995.
11. Durlak JA, DuPre EP. Implementation matters: a review of research on the influence of implementation on program outcomes and the factors affecting implementation. Am J Community Psychol 2008 Jun;41(3-4):327-350. [doi: 10.1007/s10464-008-9165-0] [Medline: 18322790]
12. Rohrbach LA, Gunning M, Sun P, Sussman S. The Project Towards No Drug Abuse (TND) dissemination trial: implementation fidelity and immediate outcomes. Prev Sci 2010 Mar;11(1):77-88 [FREE Full text] [doi: 10.1007/s11121-009-0151-z] [Medline: 19757052]
13. Domitrovich C, Greenberg M. The Study of Implementation: Current Findings From Effective Programs that Prevent Mental Disorders in School-Aged Children. Journal of Educational and Psychological Consultation 2000 Jun;11(2):193-221. [doi: 10.1207/S1532768XJEPC1102_04]
14. Durlak JA. Why Program Implementation is Important. Journal of Prevention & Intervention in the Community 1998 Jul 06;17(2):5-18. [doi: 10.1300/J005v17n02_02]
15. Ahmed N, Flisher AJ, Mathews C, Jansen S, Mukoma W, Schaalma H. Process evaluation of the teacher training for an AIDS prevention programme. Health Educ Res 2006 Oct;21(5):621-632 [FREE Full text] [doi: 10.1093/her/cyl031] [Medline: 16740671]



XSL•FO
RenderX

16. Dusenbury L, Brannigan R, Falco M, Hansen WB. A review of research on fidelity of implementation: implications for drug abuse prevention in school settings. Health Educ Res 2003 Apr;18(2):237-256 [FREE Full text] [Medline: 12729182]
17. Kramer L, Laumann G, Brunson L. Implementation and Diffusion of the Rainbows Program in Rural Communities: Implications for School-Based Prevention Programming. Journal of Educational and Psychological Consultation 2000 Mar;11(1):37-64. [doi: 10.1207/s1532768Xjepc1101_04]
18. Rohrbach LA, Graham JW, Hansen WB. Diffusion of a school-based substance abuse prevention program: predictors of program implementation. Prev Med 1993 Mar;22(2):237-260. [doi: 10.1006/pmed.1993.1020] [Medline: 8483862]
19. Han SS, Weiss B. Sustainability of teacher implementation of school-based mental health programs. J Abnorm Child Psychol 2005 Dec;33(6):665-679. [doi: 10.1007/s10802-005-7646-2] [Medline: 16328743]
20. Hofstetter H, Peters L, Meijer S, van KH, Schutte L, van Empelen P. Evaluation of the effectiveness and implementation of the sexual health program Long Live Love. European Health Psychologist 2014;16:489.
21. Wiefferink CH, Poelman J, Linthorst M, Vanwesenbeeck I, van Wijngaarden JCM, Paulussen TW. Outcomes of a systematically designed strategy for the implementation of sex education in Dutch secondary schools. Health Educ Res 2005 Jun;20(3):323-333 [FREE Full text] [doi: 10.1093/her/cyg120] [Medline: 15385429]
22. Bos V, de JD, Paulussen T. RIVM. 2010 Apr 15. Gezondheidsbevordering en preventie in het onderwijs. Stand van zaken, effectiviteit en ervaringen van GGD-en en scholen URL: http://www.rivm.nl/dsresource?objectid=rivmp:11705&type=org&disposition=inline&ns_nc=1 [WebCite Cache ID 6arxQ3vbj]
23. Zwaneveld B, Rigter H. Ruud de Moor Centrum - Open Universiteit. 2009. Over drempels naar meer ict-gebruik in het voortgezet onderwijs URL: http://www.ou.nl/documents/14300/a31352f1-9f74-4c25-8009-5919af2caff2 [accessed 2015-08-18] [WebCite Cache ID 6as30uLGd]
24. Marino G. E-coaching: Connecting learners to solutions. In: The Performance Improvement Emporium (PIE). San Diego: San Diegoan Diego State University; 2004.
25. Bartholomew K, Parcel G, Kok G, Gottlieb N, Fernandez M. Planning health promotion programs. An Intervention Mapping Approach. In: Planning health promotion programs. An Intervention Mapping Approach. San Francisco: Jossey-Bass; 2011.
26. van Bruinessen Inge Renske, van Weel-Baumgarten Evelyn M, Snippe HW, Gouw H, Zijlstra JM, van DS. Active patient participation in the development of an online intervention. JMIR Res Protoc 2014;3(4):e59 [FREE Full text] [doi: 10.2196/resprot.3695] [Medline: 25379679]
27. Makai P, Perry M, Robben SHM, Schers HJ, Heinen MM, Olde Rikkert MGM, et al. Evaluation of an eHealth intervention in chronic care for frail older people: why adherence is the first target. J Med Internet Res 2014;16(6):e156 [FREE Full text] [doi: 10.2196/jmir.3057] [Medline: 24966146]
28. Voogt CV, Poelen EAP, Kleinjan M, Lemmers Lex ACJ, Engels Rutger CME. The development of a web-based brief alcohol intervention in reducing heavy drinking among college students: an Intervention Mapping approach. Health Promot Int 2014 Dec;29(4):669-679. [doi: 10.1093/heapro/dat016] [Medline: 23525645]
29. Van de Bongardt D, Mouthaan I, Bos H. Pedagogiek. 2009. Seksuele en relationele vorming in het voortgezet onderwijs URL: http://www.pedagogiek-online.nl/index.php/pedagogiek/article/viewFile/372/371 [WebCite Cache ID 6av00uHV4]
30. Hansen WB, Bishop DC, Bryant KS. Using online components to facilitate program implementation: impact of technological enhancements to all stars on ease and quality of program delivery. Prev Sci 2009 Mar;10(1):66-75. [doi: 10.1007/s11121-008-0118-5] [Medline: 19067165]
31. Kam C, Greenberg MT, Walls CT. Examining the role of implementation quality in school-based prevention using the PATHS curriculum. Promoting Alternative THinking Skills Curriculum. Prev Sci 2003 Mar;4(1):55-63. [Medline: 12611419]
32. Mihalic SF, Fagan AA, Argamaso S. Implementing the LifeSkills Training drug prevention program: factors related to implementation fidelity. Implement Sci 2008;3:5 [FREE Full text] [doi: 10.1186/1748-5908-3-5] [Medline: 18205919]
33. Farmer‐Dougan V, Viechtbauer W, French T. Peer‐prompted Social Skills: the role of teacher consultation in student success. Educational Psychology 1999 Jun;19(2):207-219. [doi: 10.1080/0144341990190207]
34. Jones KM, Wickstrom KF, Friman PC. The effects of observational feedback on treatment integrity in school-based behavioral consultation. School Psychology Quarterly 1997;12(4):316-326. [doi: 10.1037/h0088965]
35. Mortenson B, Witt J. The use of weekly performance feedback to increase teacher implementation of a prereferral academic intervention. School Psychology Review 1998;27:613-627.
36. Noell GH, Witt JC, Gilbertson DN, Ranier DD, Freeland JT. Increasing teacher intervention implementation in general education settings through consultation and performance feedback. School Psychology Quarterly 1997;12(1):77-88. [doi: 10.1037/h0088949]
37. Witt J, Noelle G, LaFleur L, Mortenson B. Teacher use of interventions in general education: measurement and analysis of the independent variable. Journal of Applied Behaviour Analysis 1997;30:693-696. [doi: 10.1901/jaba]
38. Hesselink A, van der Sluis M, Martens M. Implementatieonderzoek Lekker fit! ResCon, research & consultancy, Amsterdam;2011. URL: http://www.rescon.nl/wp-content/uploads//2011/10/10.13.-NHS.-Lekker-fit-management-summary-OP-WEBSITE.pdf [accessed 2015-08-18] [WebCite Cache ID 6as0WL5Rk]
39. Ajzen I. The theory of planned behaviour: reactions and reflections. Psychol Health 2011 Sep;26(9):1113-1127. [doi: 10.1080/08870446.2011.613995] [Medline: 21929476]
40. Bandura A. Social foundations of thought and action: a social cognitive theory. Englewood Cliffs, N.J: Prentice-Hall; 1986.




XSL•FO
RenderX

41. Parsloe E, Leedham M. Coaching and mentoring: practical conversations to improve learning. London: Kogan Page; 2009.
42. McAlister A. Behavioral Journalism: Beyond the Marketing Model for Health Communication. American Journal of Health Promotion 1995 Jul;9(6):417-420. [doi: 10.4278/0890-1171-9.6.417]
43. Cordingley P, Bell M, Thomason S, Firth A. How do collaborative and sustained CPD and sustained but not collaborative CPD affect teaching and learning? EPPI-Centre, University of London. 2005. The impact of collaborative continuing professional development (CPD) on classroom teaching and learning URL: http://www.curee.co.uk/files/publication/1246025289/The%20impact%20of%20collaborative%20CPD%20on%20classroom%20teaching%20and%20learning.pdf [accessed 2015-08-18] [WebCite Cache ID 6aryhQwYF]
44. Cordingley P, Bell M, Rundell B, Evans D, Curtis A. How does collaborative Continuing Professional Development (CPD) for teachers of the 5-16 age range affect teaching and learning? EPPI- Centre, University of London. 2003. The impact of collaborative CPD on classroom teaching and learning URL: http://www.curee.co.uk/files/publication/1246025680/How%20does%20collaborative%20CPD%20for%20teachers%20of%20the%205-16%20age%20range%20affect%20teaching%20and%20learning.pdf [accessed 2015-08-18] [WebCite Cache ID 6aryH5TwW]
45. Zwart R, Wubbels T, Bergen T, Bolhuis S. Experienced teacher learning within the context of reciprocal peer coaching. Teachers and Teaching: Theory and Practice 2007;13:165-187. [doi: 10.1080/13540600601152520R]
46. Nielsen J. Designing Web Usability: The Practice of Simplicity. Indianapolis: New Riders Press; 1999.
47. Crutzen R, Cyr D, de Vries N. The role of user control in adherence to and knowledge gained from a website: randomized comparison between a tunneled version and a freedom-of-choice version. J Med Internet Res 2012;14(2):e45 [FREE Full text] [doi: 10.2196/jmir.1922] [Medline: 22532074]
48. Martens M, van AP, Paulussen T, Schaalma H, Brug J. Krachtvoer: process evaluation of a Dutch programme for lower vocational schools to promote healthful diet. Health Educ Res 2006 Oct;21(5):695-704 [FREE Full text] [doi: 10.1093/her/cyl082] [Medline: 16945986]
49. Dusenbury L, Brannigan R, Hansen WB, Walsh J, Falco M. Quality of implementation: developing measures crucial to understanding the diffusion of preventive interventions. Health Educ Res 2005 Jun;20(3):308-313 [FREE Full text] [doi: 10.1093/her/cyg134] [Medline: 15522898]
50. Fielden S. Literature review: coaching effectiveness- a summary. NHS. Modernisation Agency Leadership Centre Manchester: The University of Manchester; 2005. URL: http://literacy.kent.edu/coaching/information/Research/NHS_CDWPCoachingEffectiveness.pdf [accessed 2015-08-18] [WebCite Cache ID 6arzIITlr]
51. Ertmer PA, Ottenbreit-Leftwich AT. Teacher Technology Change. Journal of Research on Technology in Education 2010 Mar;42(3):255-284. [doi: 10.1080/15391523.2010.10782551]
52. Kirby DB, Laris BA, Rolleri LA. Sex and HIV education programs: their impact on sexual behaviors of young people throughout the world. J Adolesc Health 2007 Mar;40(3):206-217. [doi: 10.1016/j.jadohealth.2006.11.143] [Medline: 17321420]
53. Rogers E. Diffusion of innovations. 5th edition. New York: Free Press; 2003.
54. Pijpers A, van Montfort K, Heemstra F. Acceptatie van ICT. Theorie en een veldonderzoek onder topmanagers. Bedrijfskunde 2002;74(4):76-84.


## Abbreviations

**IM:** intervention mapping
**LLL:** Long Live Love
**MHS:** Municipal Health Services
**SRH:** sexual reproductive health


*Edited by G Eysenbach; submitted 20.08.15; peer-reviewed by J Chen, MJ Masanet, E Jerves; comments to author 31.12.15; revised version received 08.02.16; accepted 21.02.16; published 12.07.16.*

*Please cite as:*
*Schutte L, van den Borne M, Kok G, Meijer S, Mevissen FEF*
*Innovatively Supporting Teachers' Implementation of School-Based Sex Education: Developing A Web-Based Coaching Intervention From Problem to Solution*
*J Med Internet Res 2016;18(7):e136*
*URL: http://www.jmir.org/2016/7/e136/*
*doi:10.2196/jmir.5058*
*PMID:27405241*




XSL•FO
RenderX

©Lisette Schutte, Marieke van den Borne, Gerjo Kok, Suzanne Meijer, Fraukje EF Mevissen. Originally published in the Journal of Medical Internet Research (http://www.jmir.org), 12.07.2016. This is an open-access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/2.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on http://www.jmir.org/, as well as this copyright and license information must be included.



XSL•FO
RenderX

Original Paper

# Parent-Focused Childhood and Adolescent Overweight and Obesity eHealth Interventions: A Systematic Review and Meta-Analysis

Megan L Hammersley[1], MSc; Rachel A Jones[1], PhD; Anthony D Okely[1], EdD

Early Start Research Institute, Faculty of Social Sciences, University of Wollongong, Wollongong, Australia

**Corresponding Author:**
Megan L Hammersley, MSc
Early Start Research Institute
Faculty of Social Sciences
University of Wollongong
Building 21
Northfields Ave
Wollongong, 2522
Australia
Phone: 61 403233875
Fax: 61 24221 3892
Email: megan.hammersley1@uowmail.edu.au

## Abstract

**Background:** Effective broad-reach interventions to reduce childhood obesity are needed, but there is currently little consensus on the most effective approach. Parental involvement in interventions appears to be important. The use of eHealth modalities in interventions also seems to be promising. To our knowledge, there have been no previous reviews that have specifically investigated the effectiveness of parent-focused eHealth obesity interventions, a gap that this systematic review and meta-analysis intends to address.

**Objective:** The objective of this study was to review the evidence for body mass index (BMI)/BMI z-score improvements in eHealth overweight and obesity randomized controlled trials for children and adolescents, where parents or carers were an agent of change.

**Methods:** A systematic review and meta-analysis was conducted, which conforms to the Preferred Reporting Items for Systematic Reviews and Meta-analyses (PRISMA) statement. Seven databases were searched for the period January 1995 to April 2015. Primary outcome measures were BMI and/or BMI z-score at baseline and post-intervention. Secondary outcomes included diet, physical activity, and screen time. Interventions were included if they targeted parents of children and adolescents aged 0-18 years of age and used an eHealth medium such as the Internet, interactive voice response (IVR), email, social media, telemedicine, or e-learning.

**Results:** Eight studies were included, involving 1487 parent and child or adolescent dyads. A total of 3 studies were obesity prevention trials, and 5 were obesity treatment trials. None of the studies found a statistically significant difference in BMI or BMI z-score between the intervention and control groups at post-intervention, and a meta-analysis demonstrated no significant difference in the effects of parent-focused eHealth obesity interventions compared with a control on BMI/BMI z-score (Standardized Mean Difference −0.15, 95% CI −0.45 to 0.16, Z=0.94, *P*=.35). Four of seven studies that reported on dietary outcomes demonstrated significant improvements in at least 1 dietary measurement, and 1 of 6 studies that reported on physical activity outcomes demonstrated significant improvements compared with the control. The quality of the interventions was generally not high; therefore, these results should be interpreted with caution.

**Conclusion:** It is recommended that larger, longer duration, high-quality parent-focused eHealth studies are conducted, which transform successful components from face-to-face interventions into an eHealth format and target younger age groups in particular.

**Trial Registration:** PROSPERO International Prospective Register of Systematic Reviews: CRD42015019837; http://www.crd.york.ac.uk/PROSPERO/display_record.asp?ID=CRD42015019837 (Archived by WebCite at http://www.webcitation.org/6ivBHvBhq)



XSL•FO
RenderX

*(J Med Internet Res 2016;18(7):e203)* doi:10.2196/jmir.5893

**KEYWORDS**

overweight; obesity; child; adolescent; internet; web; online; computer; IVR; telemedicine; healthy lifestyle; dietary intake; physical activity

## Introduction

The escalating global challenge of childhood obesity has been well documented, with prevalence rates climbing to approximately 23% in developed countries and 13% in developing countries [1]. Childhood is a period of time where unhealthy behaviors such as consumption of energy-dense foods and beverages, physical inactivity, and sedentary behavior are established [2]. During this time, parental influence and role modeling play a key part in the development of such behaviors [3-5]. Parental involvement in childhood obesity interventions appears to be important, given that children are highly influenced by the family unit [6,7]. Recent systematic reviews and meta-analyses have investigated the effectiveness of parent-focused childhood obesity prevention and treatment interventions, with the weight of the evidence supporting the use of parent-focused interventions. A 2012 meta-analysis of weight-related behavior change interventions for 2- to 19-year olds where parents were involved resulted in greater body mass index (BMI) reductions than interventions that had optional or no parent involvement [4]. These are similar findings to 2 meta-analyses of children aged 5-12 years [8,9], whereas another meta-analysis of 2- to 18-year olds found that interventions that targeted parents had a smaller (yet still significant) effect than those that targeted children directly [10].

The lack of studies in preschool-aged children has been highlighted [11]. Of the aforementioned 2 meta-analyses that sought to include studies, which involved children from 2 years of age, one included no studies in the preschool age group and the other included only 2 studies in this age group [4,10]. A meta-analysis of parent-focused obesity prevention and treatment interventions specifically in the early childhood (0-6 years) age group demonstrated a small, yet significant combined effect in the short term, but in the long term, the combined results were not significant [2]. When the studies were looked at individually, 5 were successful in the long term, which were all commenced at preschool age. The baseline BMI of the children appeared to be a factor, as 2 of the 3 studies that were successful at both short- and long-term follow-up included only children who were overweight or obese [2].

Effective broad-reach interventions that target childhood are required; however, currently, there is little consensus on the most effective intervention approach [11]. As mentioned, interventions that target parents are effective [2,4,8]. In addition, the use of eHealth interventions also hold promise in this area, with the use of such technology in the child and adolescent age group having increased in recent years [12]. Two previous reviews have investigated the impact of technology-based overweight and obesity interventions in childhood and adolescence with some studies reporting changes in adiposity, dietary, and/or physical activity outcomes [12,13]. However, neither of these previous reviews have specifically investigated the effect of parent involvement.

This current systematic review and meta-analysis builds on previous reviews, but differs in that it is, to our knowledge, the first to measure the efficacy of eHealth interventions in improving BMI or BMI z-score in children and adolescents where parents are an agent of change. This review is of importance in determining effective broad-reach approaches to prevent and treat childhood obesity, which in the long term could potentially alter the path of childhood obesity and reduce the progression into adult life. The review adopts a broader definition of eHealth than 1 of the previous reviews and includes interventions using the Internet, IVR (computerized voice prompts over the telephone, which participants respond to via the telephone keypad), social media (Facebook, Twitter, and so forth), mobile health (such as mobile phone apps), telemedicine (using video conferencing), email, and e-learning. The objective of this current systematic review and meta-analysis was to determine whether eHealth childhood and adolescent overweight and obesity interventions, where parents or carers are the agents of change, improved BMI and/or BMI z-scores.

## Methods

The protocol for this systematic review and meta-analysis was registered in advance with the PROSPERO international prospective register of systematic reviews (registration number CRD42015019837) and conforms to the Preferred Reporting Items for Systematic Reviews and Meta-analyses (PRISMA) statement [14].

### Eligibility Criteria

#### *Type of Studies*

Randomized controlled trials investigating the effect of eHealth interventions on weight of children and adolescents, where parents or carers were an agent of change, were considered for this systematic review and meta-analysis. Studies were excluded if participants had special needs or had a condition where physical activity was restricted or if they required a special diet. Studies not published in English were also excluded.

#### *Type of Participants*

eHealth studies targeting obesity prevention or treatment for children and adolescents aged 0-18 years, where parents or carers were agents of change, were considered. The parent or carer being an agent of change was defined as the parent or carer having an active role in the intervention and being responsible for implementing change.

#### *Types of Interventions*

Interventions investigating the effect of eHealth on BMI were considered for inclusion. No restrictions were placed on the



XSL•FO
RenderX

type of setting, provided that the parent or carer was an agent of change.

### *Types of Outcome Measures*

Primary outcome measures were BMI and/or BMI z-score at baseline and post-intervention. Secondary outcomes included body fat, waist-to-hip ratio, and improvements to dietary intake, physical activity, sedentary behavior, screen time, biomedical indicators (such as blood pressure and cholesterol), knowledge, and self-efficacy.

## Search Strategy

The electronic databases of A+ Education, CINAHL, ProQuest Central, PsycINFO, Scopus, SPORTDiscus, and Web of Science were searched with a limitation date of January 1995 to April 2015 using predetermined search terms (see Multimedia Appendix 1). Pre-1995 articles were not included as it was thought that any interventions at this early stage would be exceedingly basic. In addition, the reference lists of relevant articles were scanned.

## Study Selection

After the database searches, 1 author (MH) removed duplicates and screened the titles of the articles, and relevant articles were shortlisted. A second author (RJ) then checked the decisions made. The abstracts of the remaining articles were then screened (by MH), and a second shortlist was derived and checked by a second author (RJ). The full text of the remaining articles was retrieved and read by author one to create a final shortlist. The shortlisted articles were then viewed by the second author (RJ). Any differences were discussed, and a decision was made by consensus. Where a decision could not be reached, a third author (AO) reviewed the papers to make a final decision.

## Data Collection Process

One review author (MH) independently extracted the data from the included studies. Contact was made via email with the author of 1 paper to request additional data on BMI at a time point during the study, which was used in the meta-analysis and systematic review.

## Risk of Bias in Individual Studies

Two reviewers (AO and MH) independently assessed risk of bias using a checklist adapted from the Consolidated Standards of Reporting Trials statement (see Table 1) [15]. In line with the recommendations of the PRISMA statement, each of the items on the checklist was evaluated separately rather than an overall score being assigned. Each item was given a + or − according to whether the item was described adequately in the article (+) or not adequately described or not present (−). Any differences were discussed, and a decision was made by consensus.

**Table 1.** Risk of bias checklist.

| Item | Description |
|---|---|
| A | Key baseline characteristics are presented separately for treatment groups (age, gender, and body mass index—BMI), baseline outcomes were statistically tested, and results of tests were provided |
| B | Randomization procedure clearly and explicitly described and adequately carried out (generation of allocation sequence, allocation of concealment, and implementation) |
| C | Valid measurement of BMI (at minimum, standardized method used to measure height and weight and to calculate BMI are described) |
| D | Dropout described and ≤20% for <6-month follow-up or ≤30% for ≥6-month follow-up |
| E | Blinded outcome assessment (positive when those responsible for assessing BMI were blinded to the group allocation of individual participants) |
| F | Intention-to-treat analysis for BMI outcome(s) (participants analyzed in group they were originally allocated to and participants were not excluded from analyses because of noncompliance to treatment or because of missing data |
| G | Covariates accounted for in analyses (eg, baseline score, group or cluster, and other covariates when appropriate for age or gender) |
| H | Summary results for each group and adjusted scores presented (adjusted difference between groups and CI) |
| I | Power calculation reported, and the study was adequately powered to detect hypothesized relationships |

## Synthesis of Results

Extracted data were first described in a narrative manner. Studies that reported BMI or BMI z-score results as change scores or baseline and final values; standard deviation (SD), standard error (SE), or CIs; and the number of participants were included in a meta-analysis. Mean change was calculated where required, and SDs were calculated from SE or CI where SD was not reported [16]. Where the final SD value was missing, this value was imputed from baseline SD [16]. Missing SD change values were calculated using an imputed correlation coefficient [16].

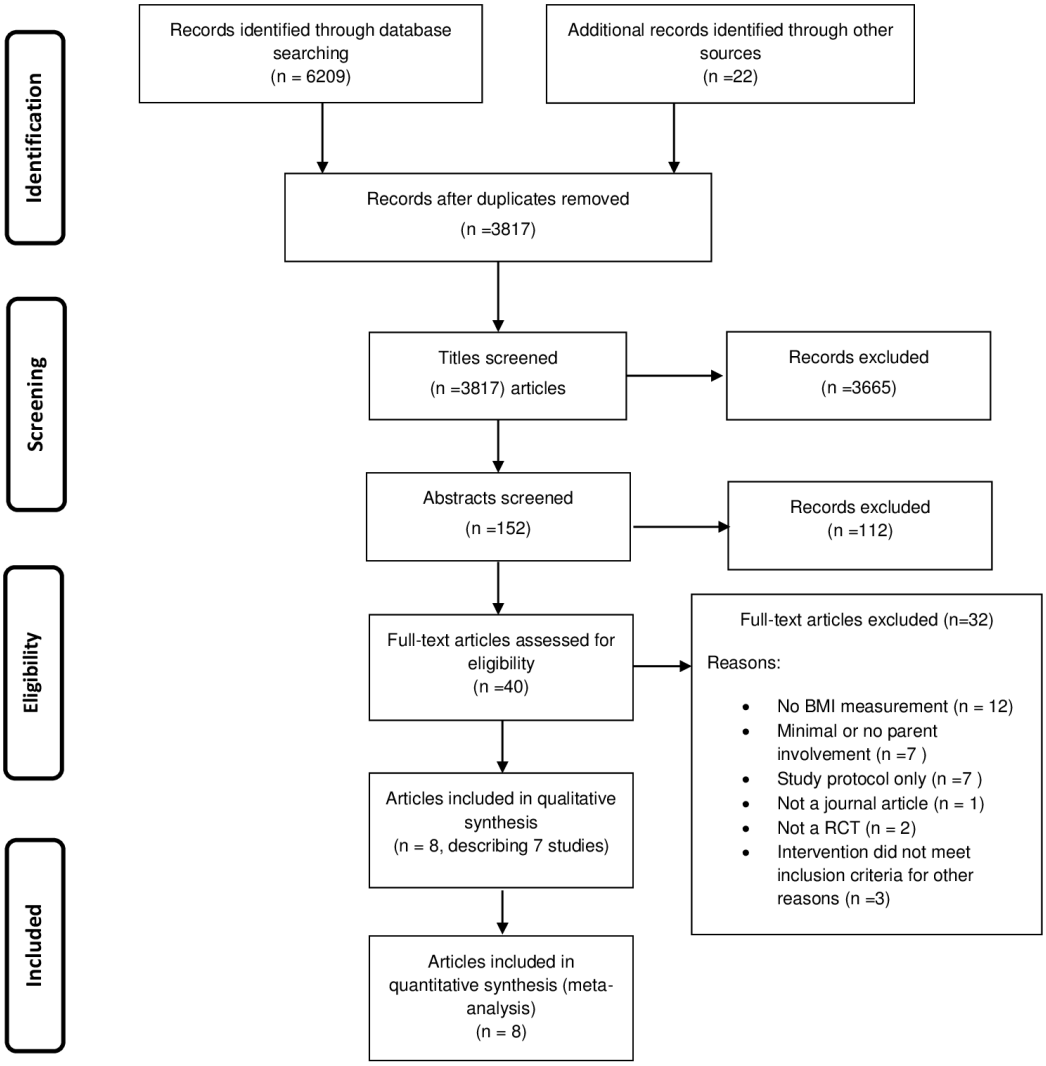
Where a study had 2 eHealth intervention arms, the number of participants in the control group was divided by 2 to ensure that participants were not counted more than once in the analysis. Heterogeneity was assessed via I2 index test. The meta-analysis was conducted with reported or calculated change scores for the data collection point closest to the end of the intervention. One study was reported across 2 articles [17,18], and the time points in both these articles were used (baseline to 6 months and 6 months to 2 years—which was calculated from the available data). To enable either BMI or BMI z-score to be included in the same meta-analysis, standardized mean difference (SMD) was used. Where a study reported both BMI and BMI z-score, BMI was used. One study involved a day camp before the implementation of the eHealth intervention, and therefore, the post-camp BMI measures were used as baseline measures for the purpose of the meta-analysis to isolate



XSL•FO
RenderX

this component [19]. A random effects model was applied to the analysis given the heterogeneity across the studies [16]. Analysis was conducted using Review Manager (RevMan: computer program) version 5.3; Copenhagen: The Nordic Cochrane Centre, The Cochrane Collaboration, 2014.

**Figure 1.** Study selection flow diagram.




## Results

### Study Selection

From the 3817 papers that were initially identified, 8 papers describing 7 separate studies met the inclusion criteria (Figure 1).

### Description of Studies

Table 2 outlines the characteristics of the studies meeting the inclusion criteria; 7 studies were conducted in the past 10 years, and only 1 study was conducted outside the United States (in France) [22]. There were a total 1487 dyads participating in the included 8 studies (range 35-1013 dyads). A range of cultural or ethnic groups participated in studies, including African American (with 3 studies including only African American participants [17-19]), Latino [20], Chinese American (1 study included only Chinese-American participants [21]), and French [22].In total, 5 studies were overweight or obesity treatment interventions [17,18,20,23,24] and 3 studies overweight prevention interventions [19,21,22]. The gender proportions of the child or adolescent participants were 47.21% male and 52.79% female. Two of the studies included only girls [17,18]. Parent gender was reported in only 1 study [24], where 96% were female. In total, 3 studies involved children (range 7-10 years) [19,22,23], 3 studies involved adolescents (range 11-15 years) [17,18,21], and 2 studies included both children and adolescents (range 5-12 years) [20,24]. The length of the interventions ranged from 8 weeks to 2 years, with 4 studies being ≤12 weeks [19-21,24], 3 studies being ≤8 months [17,22,23] and 1 study being 2 years in duration [18]. Only 1 study collected follow-up data to assess maintenance of changes in the months after the completion of the intervention [21]. Retention rates were reported in 7 studies, and the average retention rate was 80% ± 6.3 (ranging from 70% to 93%) [17,18,20-24].



XSL•FO
RenderX

### Description of Interventions

Two of the studies had 3 study arms [20,22], and the remaining 6 studies had 2 study arms. Five studies used an Internet intervention [17-19,21,22], 2 used IVR [20,24], and 1 used telemedicine [23]. Of the Internet interventions, 1 used Internet only [21], and others used the Internet in combination with face-to-face counseling [17,18], telephone counseling, and nutrition lessons [22] or a camp [19]. The focus of behavior change differed between studies, with one focusing on diet, physical activity, and screen time [20]; 6 focusing on diet and physical activity [17-19,21-23] and 1 focusing on diet and screen time [24].

A theoretical framework underpinned 4 of the studies, 2 were underpinned by Social Cognitive Theory [19,24], 1 reported using a combination of trans-theoretical model and social cognitive theory [21], and 1 reported using social-ecological theory [20]. Studies varied in the level of detail that they provided regarding how the theory was used in the design of the intervention.

The level of parental involvement varied among studies. In 1 study, only the parents participated in the intervention (children were involved only at the data collection stages) [20]. In the remaining 7 studies, the parent and the child or adolescent both had active involvement in the intervention, either the child or adolescent participated in the eHealth activities with the parent together or there were separate components designed specifically for the parent and the child or adolescent [17-19,21-24].

Studies used differing measures of adiposity, with most using multiple measures. Six studies used BMI [17-19,21,22,24], 4 studies used BMI z-score [20,22-24], 4 studies used BMI percentile [17,18,23,24], 3 used body fat (measured by DEXA [17-19], and 1 study used waist-to-hip ratio [21]. Other measures included dietary intake (measured by food frequency questionnaire [17,18,20,24], 24-hour recall [17-19,23], or food records [21,22]) physical activity (measured by questionnaire [17-20,22] or accelerometer [19,21,23]), and screen time (measured by questionnaire [20,24]).

Three of the studies reported on the effect of higher usage of the interventions. One IVR study reported that participants who completed more calls significantly decreased their BMI z-score compared with the control group [20], whereas another IVR study reported that participants who were high IVR users demonstrated a significant reduction in BMI and BMI z-score compared with low IVR users [24]. One of the Internet studies [17] reported that change in percentage body fat was negatively correlated with use of an email facility to counselors, performance on quizzes, and use of an Internet weight monitoring function.

### Risk of Bias Within Studies

Table 3 summarizes the results of the risk of bias assessment for all included studies. Of the 8 studies, 6 reported key baseline characteristics separately for each study arm, and the results of statistical tests were provided. Seven studies reported an acceptable dropout rate (≤20% for <6-month follow-up or ≤30% for ≥6-month follow-up), and the remaining study did not report dropout rates. Six studies used intention-to-treat analysis for BMI outcomes, 7 studies accounted for covariates in the analysis; power calculations were reported and adequate in 5 articles. Only two studies described an adequate randomization procedure and/or reported summary results for each group with adjusted scores, and none of the studies described a valid, standardized method of BMI measurement.



XSL•FO
RenderX

**Table 2.** Summary of parent-focused childhood or adolescent obesity eHealth interventions.

| Author, Year, Country | Participants | Intervention description | Parental involvement | Behaviors targeted | Variables measured | Key findings |
|---|---|---|---|---|---|---|
| Baranowski et al 2003, USA [19] | n=35, 8 years of age, girls | 4-week camp with specially designed activities, followed by 8-week behavior change Internet intervention. Control girls attended camp with usual activities and a monthly Internet program with general health information and homework. | No parent involvement in camp. Intervention, and control parents had access to a website, which covered similar topics to girls' website. | Diet (dietary fat intake, dietary fiber, water and satiety, SSB[a]), moderate to vigorous PA[b] | Demographics, body mass index (BMI), WC[c], physical maturation, body fat (DEXA), diet (2 ×24-hour recall), PA (accelerometer and qne), preferences for PA, and SSB. | For the Internet component, no significant changes to BMI were observed. No other variables were measured at the end of the camp, so the effect of the Internet intervention on variables other than BMI could be determined. |
| Chen et al 2011, USA [21] | n=54, 12-15 years of age Chinese American | Behavior change Internet program with goal setting tailored to stage of change. 8 ×weekly sessions for children. Control participants accessed a general health information Internet site. | Parents received 3 Internet sessions over 8 weeks to increase knowledge and skills. | Diet (food pyramid, meal planning, portion size), PA | Parent height and weight, child BMI, waist-to-hip ratio, blood pressure, PA (accelerometer), diet (3-day food diary), PA and nutrition knowledge (qne), dietary and PA self-efficacy. | Significantly more participants in the intervention reduced their waist-to-hip ratio than the control group (effect size= −0.01, *P*=.02). There were also significant increases in PA (effect size=12.46, *P*=.01), increases to F&V[d] intake (effect size=0.14, *P*=.001) and increased PA knowledge (effect size=0.16, *P*=.008), and nutrition knowledge (effect size=0.18, *P*=.001). |
| Davis et al 2013, USA [23] | n=58, 5-11 years of age, rural setting | 8 × weekly telemedicine delivered psychoeducational sessions covering goal setting, diet and PA, plus 6 ×monthly sessions. Control participants visited their primary care physician to discuss set topics. | Parents met in a group separately, but at the same time as the children and covered similar content. | Nutrition (stoplight diet, portion sizes, food labels, vitamins and minerals, nutrient density), energy balance, PA, screen time, and SB[e]. | Demographics, BMI z-score, diet (24-hour recall), PA (accelerometer), child behavior checklist, behavioral pediatrics feeding assessment scale. | No statistical difference in BMI z-score between groups. There was also no significant difference between groups for kilocalories or PA. |
| Estabrooks et al 2009, USA [20] | n=220, 8-12 years of age | Group A: 2 × 2-hour weekly group sessions on nutrition, PA, problem-solving, and action planning delivered by dietitian. Group B: attended group sessions plus 10 interactive voice response (IVR) follow-up sessions, involving goal-setting at end of call. Both the groups received a workbook with homework on nutrition and PA topics. Control group received workbook only. | Parent was main agent of change (children participated in data collection only) | Weight, nutrition, PA, and parenting skills. | BMI z-score, PA and SB (questionnaire—qne), F&V and SSB[a] consumption (qne), eating disorder symptoms (qne). | No significant difference in BMI z-score between groups. Significant increase in moderate-intensity physical activity in IVR group but no difference between groups. Participants completing 6-10 IVR calls significantly reduced BMI z-score compared with other groups [F(3,148)= −2.89, *P*<.01]. |



XSL•FO
RenderX

| Author, Year, Country | Participants | Intervention description | Parental involvement | Behaviors targeted | Variables measured | Key findings |
|---|---|---|---|---|---|---|
| Paineau et al 2008, France [22] | n=1013, 7-9 years of age | All intervention families accessed a website containing information, interactive components, and other functionality. They received 30-minute dietary counseling telephone calls from a dietitian monthly for 8 months after Web-based completion of questionnaires. Children received 3 nutrition lessons at school. Children and parents received monthly newsletters. Group A: advised to reduce fat and increase complex cholesterol (CHO), Group B: advised to reduce fat and sugars and increase complex CHO. Control group received only general nutrition information at the same intervals. | Families accessed website and received mobile phone calls. Parents received monthly newsletter. | Nutrition (portions, frequency of eating, meal modification, and healthier alternatives) | Demographics, BMI, BMI z-score, body fat, WC, chest circumference, knee circumference, dietary intake (total energy, fats, sugars, complex CHO, protein) (Web-based qne and dietary records), PA (qne) | No significant difference between groups in regard to BMI or other anthropometric measures. Group A: Significantly increased complex CHO intake (mean change +10.1 (6.0-14.2) 95% CI, *P*<.05). Group B: Significantly reduced sugar intake (mean change −10.0 (−13.4 to −6.6) 95% CI, *P*<.01). Both groups A and B reduced total energy (mean change A −60 (−104 to −15) 95% CI, *P*<.05, B −96 (−146 to −45) 95% CI, *P*<.01) and fat intake (mean change A −8.2 (−10.6 to −5.8) 95% CI, *P*<.01, B −8.3 (−10.8 to −5.7), 95% CI, *P*<.01) compared with control group. No difference in PA between groups. |
| Williamson et al 2005, USA [17] | n=57, 11-15 years of age, African-American girls | Behavioral website providing nutrition information and behavior modification for 6 months. Counseling provided via email. Control group had access to general noninteractive health website. 4 face-to-face sessions over 12 weeks, focused on goal setting, behavioral contracting, monitoring of progress, and problem-solving. Control group sessions were conducted by a dietitian and included general nutrition information. | Parent and adolescent participated in the face-to-face and Internet components together | Nutrition (low energy diet, F&V, PA, food monitoring) | Demographics, BMI, BMI percentile, body fat (DEXA), eating disorders, pubertal status, dietary intake (24-hour recall and FFQ), weight loss behavior scale, child dietary self-efficacy scale, PA social support, children's eating attitudes test, satisfaction with life scale, child depression inventory, Rosenberg self-esteem scale, Kansas family life satisfaction scale, symptom checklist-90 | Participants in the intervention group lost significantly more body fat (−1.12± 0.47 standard error—SE) than the control group 0.43±0.47 SE, *P*<.05) There was a significant difference in BMI change between groups (intervention −0.19 ± 0.24 SE, <0.05, control +0.65 ± 0.23 SE, *P*<.05). Participants in the intervention group significantly reduced fat intake compared with control group (FFQ) (−145.67 ± 37.67 SE, *P*<.05) |
| Williamson et al 2006, USA [18] | n=57, 11-15 years of age, African-American girls | Behavioral website providing nutrition information and behavior modification over 2 years. Counseling provided via email. Control group had access to general noninteractive health website. 4 face-to-face sessions over 12 weeks, focused on goal setting, behavioral contracting, monitoring of progress, and problem-solving. Control group sessions were conducted by a dietitian and included general nutrition information. | Parent and adolescent participated in the face-to-face and Internet components together | Nutrition (low energy diet, F&V, PA, food monitoring). | Demographics, BMI, BMI percentile, body fat (DEXA), eating disorders, pubertal status, weight loss behavior scale, website use, computer opinion survey. | At 2 years, there was no significant difference in BMI, weight, or body fat. Higher BMI percentile at baseline was associated with greater reduction in BMI percentile. Higher weight loss behavior scale score at baseline was associated with greater improvement. In regard to reported consumption of fattening foods, there was a significant difference between groups (F (1,48) =2.08, *P*<.05). |



XSL•FO
RenderX

| Author, Year, Country | Participants | Intervention description | Parental involvement | Behaviors targeted | Variables measured | Key findings |
|---|---|---|---|---|---|---|
| Wright et al 2013, USA [24] | n=50, 9-12 years of age | Parents and children individually received 12× weekly interactive voice response (IVR) telephone counseling calls, which provided education, monitoring, and counseling on managing weight and reducing screen time. Information sent via electronic health record to the child's pediatrician and used at visit 1 month after the intervention. Control participants attended the same pediatrician visit. | Received IVR calls independently to children. | Nutrition (energy, spotlight diet, healthy alternatives, cooking and shopping, eating out), and screen time | BMI, dietary intake (energy, fat, fruits, vegetables) (qne), TV viewing time (qne) | There was no significant difference between groups for BMI, BMI z-score, dietary intake or screen time. There was a significant difference in weight (−4.0 change, *P*=.001), BMI (−1.2 change, *P*=.01), and BMI z-score (−0.1 change, *P*=.04) between high users and low users. |

[a]SSB: sugar-sweetened beverages.

[b]PA: physical activity.

[c]WC: waist circumference.

[d]F&V=fruit and vegetables.

[e]SB: sedentary behavior.

**Table 3.** Risk of bias assessment in randomized controlled trials assessing BMI outcomes of parent-focused eHealth overweight and obesity interventions.

| Study | Baranowski et al 2003 | Chen et al 2011 | Davis et al 2013 | Estabrooks et al 2009 | Paineau et al 2008 | Williamson et al 2005 | Williamson et al 2006 | Wright et al 2013 |
|---|---|---|---|---|---|---|---|---|
| Baseline characteristics by group | + | − | + | + | + | + | − | + |
| Randomization described and conducted | + | − | − | − | − | − | − | + |
| Valid measurement of BMI | − | − | − | − | − | − | − | − |
| Dropout ≤20% for <6 months and ≤30% for ≥6 months | − | + | + | + | + | + | + | + |
| Blinded outcome assessment | − | − | − | − | + | − | − | − |
| Intention to treat for BMI outcomes | + | − | − | + | + | + | + | + |
| Covariates accounted for in analysis | + | − | + | + | + | + | + | + |
| Summary results + adjusted difference between groups + CI | + | + | − | − | − | − | − | − |
| Power calculation reported and power adequate | − | + | + | + | + | − | + | − |

\+ Adequately described and present.
− absent.

## Results of Individual Studies

### *Adiposity Outcomes*

None of the included studies reported a significant difference between groups for BMI, BMI z-score, BMI percentile, or percentage body fat from baseline to the end of the eHealth intervention. One study reported a significant difference in percentage body fat between groups at 6 months (−1.12 ± 0.47 SE, *P*<.05) [17]; this change was not maintained at the end of the 2-year intervention [18]. One study reported a significant difference between groups for waist-to-hip ratio from baseline to the end of the intervention (effect size = −0.01, *P*=.02) but reported no significant difference for BMI between groups [21].

### *Dietary Outcomes*

Four of the seven studies that assessed dietary intake (which were all Internet interventions) demonstrated a significant difference between groups in regard to improvement in at least 1 dietary outcome, such as fruit and vegetable intake [21], nutrition knowledge [21], total energy intake [22], fat intake [17,22] and "eating less fattening foods" [18].



XSL•FO
RenderX

### *Physical Activity Outcomes*

Of the 6 studies that assessed physical activity (which was an Internet intervention), 1 study demonstrated a significant difference between groups in objectively measured physical activity and physical activity knowledge [21].

### *Screen Time Outcomes*

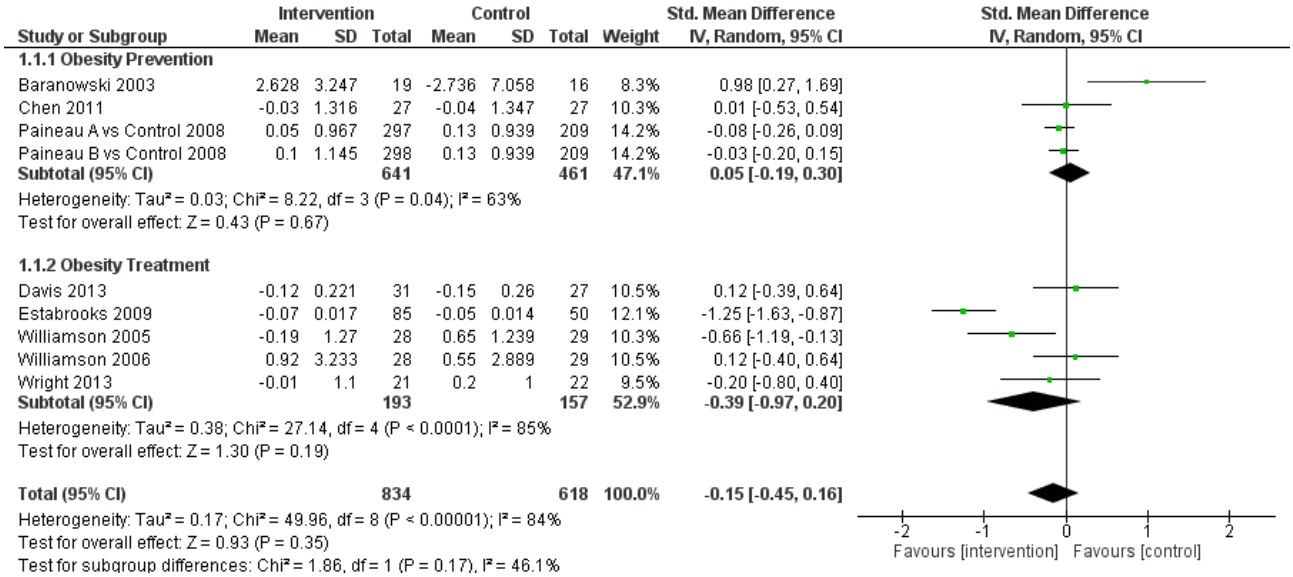
Neither of the 2 studies that assessed screen time demonstrated a significant difference between groups for screen time [22,24].

### Synthesis of Results

A meta-analysis was conducted on pooled data from 8 papers with a total of 9 study arms, which compared eHealth intervention groups with control groups. The meta-analysis results are displayed in Figure 2. The studies were found to be significantly heterogeneous (I2=84%, 95% CI: 71%-91%, *P*<.001). There was no significant difference in the effects of the eHealth interventions compared with the control groups on BMI/BMI z-score (SMD −0.15, 95% CI: −0.45 to 0.16, Z=0.93, *P*=.35). A sensitivity analysis was conducted by removing an outlying study [19], with heterogeneity decreasing slightly (I2=83%, 95% CI: 67%-91%, *P*<.001) and although the standardized mean difference moved toward favoring the intervention (−0.25, 95% CI −0.55 to 0.05), significance was not reached (Z=1.63, *P*=.10).

A sub-group analysis was conducted based on whether the study aim was obesity treatment or obesity prevention (refer to Figure 2). There was a larger effect for the obesity treatment studies (−0.39, 95% CI −0.97 to 0.20) compared with the obesity prevention studies (0.05, 95% CI −0.19 to 0.30), although this was not statistically significant. The obesity treatment studies appeared to have a higher level of heterogeneity (85%) than the obesity prevention studies (63%); however, given the small number of studies included, this should be interpreted with caution.

**Figure 2.** Effect of eHealth interventions on BMI or BMI z-score.




## Discussion

This meta-analysis and systematic review is, to our knowledge, the first to measure the effects of parent-focused eHealth childhood obesity interventions on BMI / BMI z-score. Overall, it was determined by meta-analysis that the included interventions did not result in significant improvements to BMI or BMI z-score compared with a control group. However, 4 of the 8 studies reported a significant improvement in at least 1 dietary or physical activity outcome measure.

The short duration of most of the studies may have meant there was insufficient time to detect changes in BMI or BMI z-score. The longest intervention demonstrated a significant improvement in body fat at the 6-month point [17], but this was not sustained at the end of the intervention at 2 years [18]. Maintenance of weight loss in the long term is indeed important but is a widespread challenge that has been well documented in both adult and child or adolescent age groups [11,25]. Previous parent-focused childhood or adolescent obesity systematic reviews and meta-analyses (which have not focused on eHealth) have highlighted the low proportion of studies, which have a follow-up period of >12 months [2-4,10,26], and 1 meta-analysis stated that there was a potential publication bias, meaning that it was suspected that some long-term follow-up studies with null results were not published [2]. Likewise, the lack of long-term follow-up studies has also been identified in childhood or adolescent obesity eHealth systematic reviews (which have not concentrated solely on parent-focused interventions), and it has been recommended that future interventions incorporate long-term follow-up in their design [12,13].

Maintaining engagement in eHealth interventions can be challenging [27]. The dropout rates in the current meta-analysis ranged from 12% to 29%. Previous childhood obesity eHealth systematic reviews have reported dropout rates up to 58% [12,13]. For participants that complete an eHealth intervention, the level of engagement as measured by usage rates can vary. Two of the studies in this review reported that higher usage rates resulted in more favorable BMI or BMI z-score outcomes [20,24], and 1 study found that body fat was negatively



XSL•FO
RenderX

correlated to use of an email facility to counselors, quiz results, and weight self-monitoring [17]. Conversely, lower usage rates may therefore have impacted the effectiveness of the interventions in this review. The extent of such an effect is difficult to determine as the remaining studies did not report on the differential outcomes of high users compared with low users. It is also difficult to ascertain if those who use an intervention more do so because they are more motivated, and therefore, results of comparisons between high and low users may not necessarily be indicative of the effect of the intervention itself [20]. None of the previous eHealth or parent-focused childhood obesity systematic reviews have specifically addressed the effect of usage rates on outcomes; however, it has been demonstrated in a previous systematic review on general eHealth interventions that adhere to weight-related eHealth interventions are associated with positive outcomes [28].

Most of the studies in this current review used an eHealth modality combined with face-to-face, telephone, group sessions, workbooks, or camp activities [17-20,22-24]. Only one of the interventions used eHealth as the sole mode, and interestingly, this was the only intervention to demonstrate a significant difference between groups in an anthropometric measure at the end of the intervention, with participants in the intervention group achieving a significant reduction in waist-to-hip ratio compared with the control group [21]. In regard to the studies that used other modes in addition to the eHealth mode, in most cases, it was not possible to isolate the effects of the eHealth mode, and therefore, we were not able to determine the exact effect of the eHealth component. A previous parent-focused childhood obesity systematic review found that interventions where parents received only 1 delivery mode produced better outcomes than interventions with more than 1 mode of delivery. The authors speculated that the parents may have found the intervention to be too complex when more than 1 mode was used [2], and it is possible that this may have been the case for other studies included in this current review. Previous eHealth childhood or adolescent obesity systematic reviews have discussed isolating the effects of the eHealth intervention either only briefly or not at all. Nguyen found that of the 24 studies reviewed, only 6 used eHealth as the sole mode, and 4 of these 6 studies resulted in significant improvements in BMI, BMI z-score, or obesity-related behaviors [12].

The level of parent and child or adolescent involvement in the interventions varied, but 7 of the 8 interventions involved the children or adolescents to some degree [17-19,21-24]. Only 1 of the studies delivered the intervention solely to the parent [20]. Interestingly, this was the study that was found to have the largest effect size. However, due to the small number of studies included, it is difficult to draw any conclusions from this, particularly given that the result was not statistically significant. This is similar to findings from previous parent-focused childhood or adolescent obesity systematic reviews, most of which have found that parent-focused interventions have demonstrated better outcomes than interventions where there was either no parent involvement or it was optional [4,8,9].

Three of the studies in the current review were aimed at obesity prevention and did not have being overweight or obese as an inclusion criteria. Baseline BMI or BMI z-score was therefore lower in these studies than in studies where obesity treatment was the focus, and this may have been a factor in reporting nonsignificant findings for BMI outcomes. Understandably, a subgroup analysis indicated a larger effect for obesity treatment studies compared with obesity prevention studies, which confers with a previous parent-focused childhood obesity review, which found that interventions largely aimed at obesity prevention did not significantly reduce BMI but rather prevented increases in BMI [4]. However, both these types of studies (obesity prevention and treatment) are important.

The eHealth modality used may have been a factor in the success of an intervention; however, due to the small number of studies using particular eHealth modalities (only 1 used telemedicine and 2 used interactive voice response), a subgroup analysis was not conducted. The systematic review found that 4 of the 5 Internet interventions produced positive outcomes in waist-to-hip ratio, nutrition, or physical activity measures. Internet interventions are the widest studied of eHealth modalities and have demonstrated positive effects in other recent reviews on eHealth obesity interventions [12,29].

The effectiveness of the specific content of eHealth interventions on study outcomes has not been specifically addressed in previous eHealth childhood obesity systematic reviews. In adult populations, Internet interventions with additional components such as self-monitoring, feedback, reminders, email counseling, Web-based discussion groups, Web-based lessons, text messages, social networking, or mobile phone apps have been found to be more successful in producing weight loss outcomes. Such components were used to a small extent in the studies included in this review, including monitoring [18,21,22], email counseling [18], feedback [18], and reminders [19]. The incorporation of more of these components in future eHealth childhood obesity interventions may assist in improving outcomes.

There were no interventions targeting the early childhood age group (0-5 years) in this review, and in general childhood obesity research, there has been a lack of interventions in this age group [11]. Overall, parent-focused childhood obesity interventions have been found to be effective in this age group in the short term, particularly where only 1 mode of intervention is used [2]. It has been proposed that early childhood is the ideal life stage to intervene in the course of childhood obesity as it is a time where new healthy lifestyle practices can be introduced, rather than attempting to change well-established unhealthy practices in older age groups [5]. At this stage of life, parents are usually the main influence on the nutrition and physical activity practices of their children, and therefore, the effect of parental influence is likely to be more profound than in older age groups when outside influences become more prominent [5]. Engaging parents of young children via an eHealth modality may be an appealing format for parent-focused interventions, given that parents in developed countries with children within this age group appear to be tech savvy (as suggested by a high proportion of Internet or SMART phone use) [30-33].

There were only a small number of studies found over the 20-year period included in this meta-analysis, demonstrating



XSL•FO
RenderX

that this field of study has not been well investigated, despite the dramatic advances and acceptability in technology. eHealth in childhood or adolescent obesity is only a relatively new area; a 2010 systematic review found only 21 studies, and only 11 of these were randomized controlled trials (RCTs) [12]. In this current parent-focused review, there was only 1 study found that was over 10 years old.

The quality of the interventions was generally not high, with the areas of randomization, blinded outcome assessment, valid measurement of BMI, and adjusted difference between groups either not being described or adequately carried out in a number of studies. The results should therefore be interpreted with caution due to potential bias. This is a similar finding to a previous eHealth childhood obesity review [12].

### Strengths and Limitations

The strengths of this meta-analysis and systematic review include adherence to a registered study protocol and rigorous use of the PRISMA statement. A detailed search strategy was used over several databases with a wide date range, and strict inclusion criteria were applied during the study selection process. To our knowledge, this review is the first to quantitatively measure the effects of parent-focused eHealth childhood or adolescent obesity interventions on BMI or BMI z-score. Limitations of this review include the restriction to articles published only in English, the small number of RCTs found, varying study quality, heterogeneity of the studies, inadequate power to detect an outcome in some studies due to a small number of participants, inability to isolate the effects of the eHealth component of the intervention in most studies, varying aims between studies (with some studies focusing on obesity prevention and others on obesity treatment), and all but 1 study being conducted in the United States.

In regard to the meta-analysis, as previously stated in the results, there was an outlying study that favored the control group [19]. It should be noted that this study reported a significant difference in BMI measures at baseline (with the control group having a much larger mean BMI than the intervention group), which may have influenced the results. The planned subgroup analyses comparing the type of eHealth modality used and participant age were not conducted due to the small number of studies and the wide range of ages within the individual studies making it difficult to analyze different age groups. Finally, as there were less than 10 studies in the meta-analysis, a funnel plot analysis was not conducted due to the low power of this test when there are a small number of included studies [16].

### Conclusions

This systematic review and meta-analysis found that there was no significant reduction in BMI or BMI z-score resulting from parent-focused eHealth childhood or adolescent obesity interventions compared with control. Only 1 study found a significant change in weight or adiposity measures (waist-to-hip ratio), and half of the studies demonstrated significant improvements obesity-related behaviors such as diet or physical activity compared with a control group. Only 1 study used eHealth as the sole modality, making it difficult to determine the true effect of eHealth on obesity. This review highlighted key weaknesses in the current literature: most studies were generally not of high quality, many had a short duration and lack of long-term follow-up, and many included only a small number of participants; and therefore, they may have been inadequately powered. There was an absence of studies that included children aged younger than 5 years, an age group where parental influence is probably more profound than older childhood and adolescence. It is therefore recommended that larger, high-quality studies of longer duration and longer follow-up are conducted, which transform successful components from face-to-face interventions into an eHealth format, particularly those that target younger age groups.

## Conflicts of Interest

None declared

## Multimedia Appendix 1

Search strategy.

[PDF File (Adobe PDF File), 19KB - jmir_v18i7e203_app1.pdf ]

## References

1. Ng M, Fleming T, Robinson M, Thomson B, Graetz N, Margono C, et al. Global, regional, and national prevalence of overweight and obesity in children and adults during 1980-2013: a systematic analysis for the Global Burden of Disease Study 2013. Lancet 2014 Aug 30;384(9945):766-781. [doi: 10.1016/S0140-6736(14)60460-8] [Medline: 24880830]
2. Yavuz HM, van Ijzendoorn MH, Mesman J, van der Veek S. Interventions aimed at reducing obesity in early childhood: a meta-analysis of programs that involve parents. J Child Psychol Psychiatry 2015 Jun;56(6):677-692. [doi: 10.1111/jcpp.12330] [Medline: 25292319]
3. Golley RK, Hendrie GA, Slater A, Corsini N. Interventions that involve parents to improve children's weight-related nutrition intake and activity patterns - what nutrition and activity targets and behaviour change techniques are associated with intervention effectiveness? Obes Rev 2011 Feb;12(2):114-130. [doi: 10.1111/j.1467-789X.2010.00745.x] [Medline: 20406416]



XSL•FO
RenderX

4. Niemeier BS, Hektner JM, Enger KB. Parent participation in weight-related health interventions for children and adolescents: a systematic review and meta-analysis. Prev Med 2012 Jul;55(1):3-13. [doi: 10.1016/j.ypmed.2012.04.021] [Medline: 22575353]
5. Natale RA, Messiah SE, Asfour L, Uhlhorn SB, Delamater A, Arheart KL. Role modeling as an early childhood obesity prevention strategy: effect of parents and teachers on preschool children's healthy lifestyle habits. J Dev Behav Pediatr 2014;35(6):378-387. [doi: 10.1097/DBP.0000000000000074] [Medline: 25007060]
6. Gruber KJ, Haldeman LA. Using the family to combat childhood and adult obesity. Prev Chronic Dis 2009 Jul;6(3):A106 [FREE Full text] [Medline: 19527578]
7. Ventura AK, Birch LL. Does parenting affect children's eating and weight status? Int J Behav Nutr Phys Act 2008;5:15 [FREE Full text] [doi: 10.1186/1479-5868-5-15] [Medline: 18346282]
8. Young KM, Northern JJ, Lister KM, Drummond JA, O'Brien WH. A meta-analysis of family-behavioral weight-loss treatments for children. Clin Psychol Rev 2007 Mar;27(2):240-249. [doi: 10.1016/j.cpr.2006.08.003] [Medline: 17070638]
9. Ewald H, Kirby J, Rees K, Robertson W. Parent-only interventions in the treatment of childhood obesity: a systematic review of randomized controlled trials. J Public Health (Oxf) 2014 Sep;36(3):476-489 [FREE Full text] [doi: 10.1093/pubmed/fdt108] [Medline: 24273229]
10. Peirson L, Fitzpatrick-Lewis D, Morrison K, Warren R, Usman AM, Raina P. Treatment of overweight and obesity in children and youth: a systematic review and meta-analysis. CMAJ Open 2015;3(1):E35-E46 [FREE Full text] [doi: 10.9778/cmajo.20140047] [Medline: 25844368]
11. Oude LH, Baur L, Jansen H, Shrewsbury VA, O'Malley C, Stolk RP, et al. Interventions for treating obesity in children. Cochrane Database Syst Rev 2009(1):CD001872. [doi: 10.1002/14651858.CD001872.pub2] [Medline: 19160202]
12. Nguyen B, Kornman KP, Baur LA. A review of electronic interventions for prevention and treatment of overweight and obesity in young people. Obes Rev 2011 May;12(5):e298-e314. [doi: 10.1111/j.1467-789X.2010.00830.x] [Medline: 21348921]
13. An JY, Hayman LL, Park YS, Dusaj TK, Ayres CG. Web-based weight management programs for children and adolescents: a systematic review of randomized controlled trial studies. ANS Adv Nurs Sci 2009;32(3):222-240. [doi: 10.1097/ANS.0b013e3181b0d6ef] [Medline: 19707091]
14. Moher D, Liberati A, Tetzlaff J, Altman DG. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. BMJ 2009 Jul 21;339(jul21 1):b2535. [doi: 10.1136/bmj.b2535]
15. Schulz KF, Altman DG, Moher D. CONSORT 2010 statement: updated guidelines for reporting parallel group randomized trials. Ann Intern Med 2010 Jun 1;152(11):726-732. [doi: 10.7326/0003-4819-152-11-201006010-00232] [Medline: 20335313]
16. Higgins JPT, Green S, editors. Cochrane Handbook for Systematic Reviews of Interventions Version 5.1.0. London: The Cochrane Collaboration; 2011.
17. Williamson DA, Martin PD, White MA, Newton R, Walden H, York-Crowe E, et al. Efficacy of an internet-based behavioral weight loss program for overweight adolescent African-American girls. Eat Weight Disord 2005 Sep;10(3):193-203. [Medline: 16277142]
18. Williamson DA, Walden HM, White MA, York-Crowe E, Newton RL, Alfonso A, et al. Two-year internet-based randomized controlled trial for weight loss in African-American girls. Obesity (Silver Spring) 2006 Jul;14(7):1231-1243. [doi: 10.1038/oby.2006.140] [Medline: 16899804]
19. Baranowski T, Baranowski JC, Cullen KW, Thompson DI, Nicklas T, Zakeri IE, et al. The Fun, Food, and Fitness Project (FFFP): the Baylor GEMS pilot study. Ethn Dis 2003;13(1 Suppl 1):S30-S39. [Medline: 12713209]
20. Estabrooks PA, Shoup JA, Gattshall M, Dandamudi P, Shetterly S, Xu S. Automated telephone counseling for parents of overweight children: a randomized controlled trial. Am J Prev Med 2009 Jan;36(1):35-42. [doi: 10.1016/j.amepre.2008.09.024] [Medline: 19095163]
21. Chen J, Weiss S, Heyman MB, Cooper B, Lustig RH. The efficacy of the web-based childhood obesity prevention program in Chinese American adolescents (Web ABC study). J Adolesc Health 2011 Aug;49(2):148-154 [FREE Full text] [doi: 10.1016/j.jadohealth.2010.11.243] [Medline: 21783046]
22. Paineau DL, Beaufils F, Boulier A, Cassuto D, Chwalow J, Combris P, et al. Family dietary coaching to improve nutritional intakes and body weight control: a randomized controlled trial. Arch Pediatr Adolesc Med 2008 Jan;162(1):34-43. [doi: 10.1001/archpediatrics.2007.2] [Medline: 18180410]
23. Davis AM, Sampilo M, Gallagher KS, Landrum Y, Malone B. Treating rural pediatric obesity through telemedicine: outcomes from a small randomized controlled trial. J Pediatr Psychol 2013 Oct;38(9):932-943 [FREE Full text] [doi: 10.1093/jpepsy/jst005] [Medline: 23428652]
24. Wright JA, Phillips BD, Watson BL, Newby PK, Norman GJ, Adams WG. Randomized trial of a family-based, automated, conversational obesity treatment program for underserved populations. Obesity (Silver Spring) 2013 Sep;21(9):E369-E378 [FREE Full text] [doi: 10.1002/oby.20388] [Medline: 23512915]
25. Jones R, Wells M, Okely A, Lockyer L, Walton K. Is an online healthy lifestyles program acceptable for parents of preschool children? Nutrition & Dietetics 2011;68(2):149-154 [FREE Full text] [doi: 10.1111/j.1747-0080.2011.01514.x]




XSL•FO
RenderX

26. Sung-Chan P, Sung YW, Zhao X, Brownson RC. Family-based models for childhood-obesity intervention: a systematic review of randomized controlled trials. Obes Rev 2013 Apr;14(4):265-278. [doi: 10.1111/obr.12000] [Medline: 23136914]
27. Glasgow RE. eHealth evaluation and dissemination research. Am J Prev Med 2007 May;32(5 Suppl):S119-S126. [doi: 10.1016/j.amepre.2007.01.023] [Medline: 17466816]
28. Donkin L, Christensen H, Naismith SL, Neal B, Hickie IB, Glozier N. A systematic review of the impact of adherence on the effectiveness of e-therapies. J Med Internet Res 2011;13(3):e52 [FREE Full text] [doi: 10.2196/jmir.1772] [Medline: 21821503]
29. Hutchesson MJ, Rollo ME, Krukowski R, Ells L, Harvey J, Morgan PJ, et al. eHealth interventions for the prevention and treatment of overweight and obesity in adults: a systematic review with meta-analysis. Obes Rev 2015 May;16(5):376-392. [doi: 10.1111/obr.12268] [Medline: 25753009]
30. File T, Ryan C. Computer and Internet Use in the United States. Washington, DC: US Census Bureau; 2014. URL: https://www.census.gov/history/pdf/2013comp-internet.pdf [accessed 2016-02-24] [WebCite Cache ID 6fWZin79S]
31. Australian Bureau of Statistics. Household Use of Information Technology, Australia, 2014-15. URL: http://www.abs.gov.au/ausstats/abs@.nsf/mf/8146.0 [accessed 2016-02-24] [WebCite Cache ID 6fWa7uqAd]
32. Pew Research Center. Washington DC: Pew Research Center; 2015. The Smartphone Difference URL: http://www.pewinternet.org/2015/04/01/us-smartphone-use-in-2015/ [accessed 2016-07-15] [WebCite Cache ID 6izfECl2F]
33. Australian Communications and Media Authority. Melbourne: Australian Communications and Media Authority Communications Report 2013-14 URL: http://www.acma.gov.au/~/media/Research%20and%20Analysis/Publication/Comms%20Report%202013%2014/PDF/Communications%20report%20201314_LOW-RES%20FOR%20WEB%20pdf.pdf [accessed 2016-02-25] [WebCite Cache ID 6fXvLgXtG]


## Abbreviations

**BMI:** body mass index
**BMI z-score:** body mass index z-score
**CHO:** carbohydrate
**CI:** confidence interval
**FFQ:** food frequency questionnaire
**F&V:** fruit and vegetable
**IVR:** interactive voice response
**PA:** physical activity
**Qne:** questionnaire
**RCTs:** randomized controlled trials
**SSB:** sugar-sweetened beverages
**SD:** standard deviation
**SE:** standard error
**SMD:** Standardized mean difference
**WC:** waist circumference


*Edited by G Eysenbach; submitted 21.04.16; peer-reviewed by L Korkalo, M Amini; comments to author 01.06.16; revised version received 14.06.16; accepted 17.06.16; published 21.07.16.*

*Please cite as:*
*Hammersley ML, Jones RA, Okely AD*
*Parent-Focused Childhood and Adolescent Overweight and Obesity eHealth Interventions: A Systematic Review and Meta-Analysis*
*J Med Internet Res 2016;18(7):e203*
*URL: http://www.jmir.org/2016/7/e203/*
*doi:10.2196/jmir.5893*
*PMID:27443862*



©Megan L Hammersley, Rachel A Jones, Anthony D Okely. Originally published in the Journal of Medical Internet Research (http://www.jmir.org), 21.07.2016. This is an open-access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/2.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on http://www.jmir.org/, as well as this copyright and license information must be included.




XSL•FO
RenderX

Review

# Web-Based Interventions for Chronic Back Pain: A Systematic Review

Shashank Garg[1], BSc, MSc, MD; Divya Garg[1], MD; Tanvir C Turin[1], MBBS, MS, PhD; M Faruq U Chowdhury[1], MBBS, MPH, MSc, FCGP

University of Calgary, Department of Family Medicine, University of Calgary, Calgary, AB, Canada

**Corresponding Author:**
Shashank Garg, BSc, MSc, MD
Department of Family Medicine
University of Calgary
Room G012, Health Sciences Center, 3330 Hospital Drive Northwest,
Calgary, AB, T2N 4N1
Canada
Phone: 1 403 210 7199
Fax: 1 403 210 9180
Email: sgarg@ucalgary.ca

## Abstract

**Background:** Chronic low back pain is one of the most common presenting complaints to a physician's office. Treatment is often challenging and recovery depends on various factors, often resulting in significant investments of time and resources.

**Objective:** The aim of this review is to determine which Web-based interventions aimed at chronic low back pain are of benefit to patients.

**Methods:** Randomized controlled trials (RCTs) studying Web-based interventions directed at adults with chronic low back pain were included. Retrospective studies, narrative reviews, nonrandomized trials, and observational studies were excluded. Electronic databases and bibliographies were searched.

**Results:** In total, nine unique RCTs were identified (total participants=1796). The number of patients randomized in each trial ranged from 51 to 580. Four trials studied online cognitive behavioral therapy (CBT) and five trials studied other Web-based interventions with interactive features. Empowerment/control was improved in six studies. Use of CBT was associated with reduced catastrophization among patients. Mixed results were reported with regards to reduction in pain levels and disability, although some studies showed promise in reducing disability in the short term. One study that measured health care utilization reported reduced utilization with the use of moderated email discussion.

**Conclusions:** Limited data are available regarding effective Web-based interventions to improve outcomes for patients with chronic low back pain. Nine RCTs with small sample sizes were identified in this review. Online CBT appears to show some promise in terms of reducing catastrophization and improving patient attitudes. Further research in this area with larger-scale studies focusing on appropriate outcomes appears to be a priority.

***(J Med Internet Res 2016;18(7):e139)*** doi:10.2196/jmir.4932

**KEYWORDS**

Internet; chronic back pain; Web-based interventions; systematic review; cognitive behavioral therapy; empowerment; disability

## Introduction

Low back pain is one of the most common presenting complaints in physicians' offices in North America [1]. Annual incidence of this condition in adults has been estimated to be between 10% and 15% worldwide [1]. The 3-month prevalence of low back and/or neck pain has been reported to be as high as 31% in the Unites States [2]. In addition to affecting the patient's physical and psychological well-being, there are many other ways this condition impacts the population's health and society in general. Back pain is a common cause of disability, absence from work and loss of productivity [3]. Back pain has significant economic repercussions including loss of productivity, morbidity, and costs to the health care system [4]. For example, Americans spend at least US $50 billion per year on low back pain [5]. Further, multiple studies have shown that absence from work affects patient's well-being negatively and an increased



XSL•FO
RenderX

length of absence makes it less likely that the individual will return to work [6,7].

Although the prognosis for low back pain remains good if the pain resolves in the acute phase (less than 3 months), individuals unable to do so may face a slow recovery at significant cost to self and the health care system [1]. Researchers have demonstrated that the treatment of back pain is complex [8]. This is because etiology may be multifaceted. Patient factors include age, presence of chronic disease, comorbidities, obesity, and sedentary lifestyle. Environmental factors can include work duties that require tasks such as heavy lifting, ergonomics, and others. Research on the effectiveness of rehabilitation interventions shows mixed results. A recent systematic review [9] reports insufficient data to draw conclusions about the effectiveness of specific interventions including back schools, massage, and patient education.

Due to the complex nature of chronic low back pain, effective treatment may include use of a multidisciplinary team (MDT). A MDT may be composed of a number of professionals, including a kinesiologist, physiotherapist, psychologist, occupational therapist, and pharmacist. A recent review reports that intensive multidisciplinary rehabilitation improves function in chronic back pain [8]. However, intensive daily rehabilitation for periods of up to 6 weeks [8] would require a significant commitment on behalf of the patient and at significant financial cost. Many patients have several barriers to access health professionals including lack of time, financial coverage, and lack of understanding of their role. Chronic pain is also known to have negative effects on the patient's propensity for "self-management" of their chronic condition. There is a need for treatment approaches that are easily accessible, cost-effective, and reduce the effort required on the part of patients.

Recently, there has been some interest in using the Internet as a channel to offer interventions to treat chronic low back pain. This has several advantages. Some of the barriers that apply to face-to-face meetings with medical professionals may be ameliorated through the Internet. For example, patients can use online resources at their own convenience and may be able to reduce their health care-related costs. It is possible that Web-based interventions may lead to patient empowerment by supporting ownership over their health thereby encouraging patients to be more proactive about the treatment, maintenance, and follow-up of their condition.

The purpose of this review is to summarize randomized controlled trials (RCTs) that assess the effectiveness of Web-based interventions to support patients with chronic low back pain.

## Methods

### Data Sources and Searches

Electronic databases were searched for relevant citations between January 2000 and September 2014, including MEDLINE, PUBMED, and EMBASE. Search terms included: "Internet based," "Internet-based," "Internet Delivered," "Internet-Delivered," "Web based," "Web-based," "World Wide Web," "Online," "Telemedicine," "Tele-medicine," "Email," "E-mail," "Mobile," "Phone," "Smartphone," "Tablet," "intervention," "treatment," "therapy," "communication," "counseling," "education," "educational," "instruction," "management," "self management," "chronic," "recurrent," "duration greater than 3 months," "low back pain," "low-back pain," "lower back pain," "lumbar," "lumbosacral region," "mechanical," "degenerative disc disease," "sciatica," "myofascial back pain," "nonspecific back pain," and "adult". Publication types and study designs of interest including systematic reviews, meta-analyses, practice guidelines, RCTs, and controlled clinical trials. Bibliographies of eligible articles were also searched for relevant studies. Selected journals were also searched individually for any relevant publications.

### Study Selection

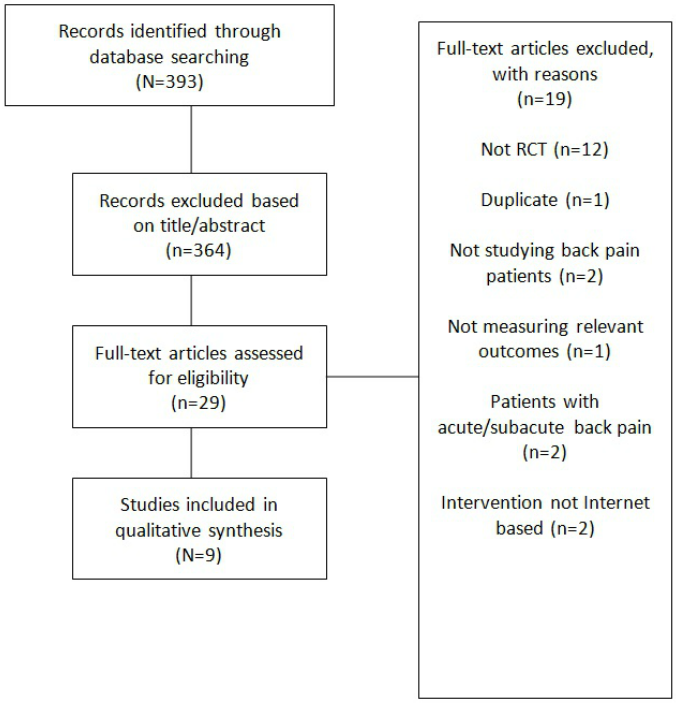
Articles were eligible for inclusion in this review if they were RCTs studying Web-based interventions directed at adults with chronic low back pain. Retrospective studies, narrative reviews, nonrandomized trials, and observational studies were excluded. However, references listed in these publications were reviewed to look for any studies that may match inclusion criteria for this review. Trials including children or trials including patients with acute pain were also excluded. RCTs studying interventions aimed at prophylaxis or other types of chronic pain were excluded. Studies published in languages other than English were excluded.

After the literature search identified potentially relevant articles, the articles were screened based on titles and abstract. Articles were excluded if they were not RCTs, the patient population was unsuitable for this review, the intervention was not Internet-based, or for other reasons (Figure 1). After this stage, the full text for the remaining articles was reviewed and nine were included for the purposes of this review.



XSL•FO
RenderX

**Figure 1.** PRISMA flow diagram.




### Data Extraction

Citations identified by the literature search strategy were screened for eligibility by two of the authors (SG, DG) and discrepancies were resolved using the opinion of the other authors. Information regarding the patient characteristics, intervention, duration, study characteristics, study design, and outcome measures was extracted from each eligible trial by one author (SG) and then reviewed independently by the other authors. Information required to assess the characteristics of studies was reviewed, including method of randomization, whether statistical analyses were performed by intention-to-treat, and allocation concealment [10-18].

## Results

### Literature Search Results

The search revealed nine RCTs published between 2002 and 2014. Table 1 describes the characteristics of each study; six of the studies performed intention-to-treat analysis, three of the studies did not describe randomization, and allocation concealment was documented in only four of the published studies.

Trial design and details regarding the interventions used in the studies are presented in Table 2. Studies randomized 51 to 580 participants [10-18]. Study durations lasted from 6 weeks to 1 year. Three of the studies were waitlist controlled.

Patient characteristics including demographics are listed in Table 3. The majority of participants in the studies using online CBT were females. The mean age of participants in the studies ranged between 42 and 52 years.

A variety of diverse outcome measures were used; outcome measures used by each study are available in Table 4. Studies using CBT reported catastrophization as an outcome measure. Most studies reported patient empowerment and pain levels as outcome measures. Disability was reported by only five studies. Only one study assessed impact of intervention on health care utilization.



XSL•FO
RenderX

**Table 1.** Study characteristics (N=9).

| Study | Year | Randomization | Intention-to-treat analysis | Allocation | Lost to follow-up, n (%) |
|---|---|---|---|---|---|
| Lorig et al [10] | 2002 | Not described | Performed by using last known data | Unclear | 159 (27.4) |
| Burhman et al [11] | 2004 | Not described | Not performed | Unconcealed | 5 (9) |
| Chaiuzzi et al [12] | 2010 | Adaptive/stratified randomization | Yes | Unclear | 10 (5) |
| Burhman et al [13] | 2011 | Webpage (random.org) | Yes | Performed through webpage | 4 (7.4) |
| Carpenter et al [14] | 2012 | Random number table | Not performed | Unclear | 23 (16.3) |
| Moessner et al [15] | 2012 | Not described | Yes | Unclear | 4 (5.3) |
| Krein et al [16] | 2013 | Random number generator | Yes | Assignment of participants through automated email message | 19 (8.2) |
| Riva et al [17] | 2014 | Random number generator (permuted block randomization method) | Yes | No face-to-face contact; no identifying information linked to patient assessment | 0 (0) |
| Weymann et al [18] | 2015 | Simple computerized randomization procedure | Yes | Concealed random allocation automatically performed using software | 180 (47) |



XSL•FO
RenderX

**Table 2.** Trial design of included studies.

| Study | Patients randomized | Intervention | Control | Duration | Measurement Time |
|---|---|---|---|---|---|
| Lorig et al [10] | 580 | Moderated email discussion group; back pain help book; videotape | Control group received usual care | 1 year | Baseline, 6 and 12 months |
| Burhman et al [11] | 56 | Web-based multimodal pain management program (CBT, stretching and exercise); weekly submission of pain diaries; weekly telephone support | Waitlist | 8 weeks | Baseline, 8-weeks and 3-months postintervention |
| Chiauzzi et al [12] | 228 | painACTION back pain website based on CBT and chronic pain management principles that provided tailored information to participants logging in twice weekly | Control group received copy of back pain help book | 6 months | Baseline, 1, 3, and 6 months |
| Burhman et al [13] | 54 | Web-based multimodal pain management program based on CBT; no weekly telephone support | Waitlist | 12 weeks | Baseline and 12 weeks |
| Carpenter et al [14] | 141 | Web-based wellness workbook | Waitlist | 6 weeks | Baseline, 3 and 6 weeks |
| Moessner et al [15] | 75 | Intervention consisted of: individualized self-monitoring module, moderated Internet-based chat | Treatment as usual | 15 weeks | Baseline, 115 and 202 days |
| Krein et al [16] | 229 | Intervention: pedometer with access to uploaded personal walking data, walking goals, feedback, participation in e-community | Enhanced usual care group also received pedometers but no access to walking goals or feedback | 12 months | Baseline, 6 and 12 months |
| Riva et al [17] | 51 | RCT with two arms: intervention and control group | Intervention group received access to back pain management website with interactive features (virtual gym, action plan, testimonials, quiz game); control group also used website, but no interactivity | 8 weeks | Baseline, 4 and 8 weeks |
| Weymann et al [18] | 382 (chronic low back pain) | Web-based information system for patients which was tailored for individual needs and dialog based | Access to information through website without tailoring or use of dialogs | 12 weeks | Baseline, first visit, and 3 months |



XSL•FO
RenderX

**Table 3.** Patient characteristics of included studies.

| Study | Patients randomized | Demographics | Inclusion criteria | Exclusion criteria |
|---|---|---|---|---|
| Lorig et al [10] | 580 | Control group: 61% male, mean age 45 years; intervention group: 62% male, mean age 46 years | One outpatient visit for back pain within last year | Continuous back pain for >90 days causing major activity intolerance; no physician visits for back pain in past year; receiving disability payments; red flag symptoms; planned back surgery; back pain due to systemic illness; pregnancy; unable to understand English |
| Burhman et al [11] | 56 | 62.5% female; mean age 44.6 years (SD 10.4) | Age 18-65 years; access to Internet; previous contact with physician; lumbar/thoracic/cervical back pain; chronic pain ≥3 months | Wheelchair bound; planned surgery; cardiovascular disease |
| Chiauzzi et al [12] | 228 | 67% female; mean age 46.14 years (SD 11.99) | Presence of back pain for ≥10 days, for ≥3 consecutive months; spinal origin of pain; English language fluency | Nonspinal medical or systemic conditions that explain the back pain; cervical pain without low back pain; psychiatric hospitalization within past year |
| Burhman et al [13] | 54 | 68.5% female; mean age 43.2 years (SD 9.8) | Access to Internet; chronic pain ≥3 months duration | Planned surgery; wheelchair bound; cardiovascular disease |
| Carpenter et al [14] | 141 | 83% female; mean age 42.5 years (SD 10.3) | Non-cancer-related back pain; duration ≥6 months; mean pain rating >4; access to Internet; | Age <40 years (applied after start of study); CBT within past 3 years; pain duration <6 months |
| Moessner et al [15] | 75 | Control group: 54.3% female, mean age 46.6 years (SD 7.7); intervention group: 57.5% female mean age 45.2 years (SD 10.2) | Age >18 years; prior multidisciplinary treatment for 1 week | Cancer-related pain; insufficient Language skills; treatment duration <1 week |
| Krein et al [16] | 229 | Control group: 86% male, mean age 51.9 years (SD 12.8); intervention group: 89% male, mean age 51.2 years (SD 12.5) | Persistent back pain; ≥3 months; self-reported sedentary lifestyle (<150 min of physical activity per week); Internet access | Inability to walk one block; pregnancy |
| Riva et al [17] | 51 | Control group: 50% female, mean age 51 years (SD 14.1); intervention group: 51.9% female, mean age 44 years (SD 13.6) | Age >18 years; back pain >3 months; Italian native speakers | Concurrent involvement in other study |
| Weymann et al [18] | 382 (chronic low back pain) | Control group: 59.1% female, mean age 52.7 years (SD 13); intervention group: 58.5% female, mean age 52.2 years (SD 13.1) | Age >18 years; chronic back pain defined as pain almost every day for period >12 weeks; diabetes type 2 | Age <18 years; duration of pain <12 weeks; lack of Internet access |



XSL•FO
RenderX

**Table 4.** Outcomes of included studies.

| Study | Outcome measures[a] | Catastrophization[b] | Empowerment/Control[b] | Pain[b] | Disability[b] |
|---|---|---|---|---|---|
| Lorig et al [10] | Pain (VNS); disability (RMQ); role function; health distress (MOS); health care utilization | NA | Increase | Increase | Increase |
| Burhman et al [11] | CSQ; MPI; PAIRS; HADS; pain diary; treatment credibility; satisfaction with treatment format | Increase | Increase | No effect | NA |
| Chaiuzzi et al [12] | BPI; ODQ; DASS; PGIC; CPCI-42; PCS; PSEQ; FABQ | Increase | Increase | No effect | No effect |
| Burhman et al [13] | CSQ; MPI; PAIRS; QOLI | Increase | No effect | No effect | NA |
| Carpenter et al [14] | Primary: SOPA; others: FABQ, NMRS, PCS, RMQ, SES | Increase | Increase | No effect | No effect |
| Moessner et al [15] | Pain intensity (NRS); SF-36; RMQ; KPD-38; Secondary: HADS (anxiety,; depression), general psychologic impairment | NA | NA | Increase | Increase |
| Krein et al [16] | Primary: RMQ, MOS; others: pain intensity, Fear-Avoidance Beliefs Questionnaire physical activity subscale | NA | Increase | No effect | Increase (6-month assessment); no effect (12-month assessment); |
| Riva et al [17] | Empowerment (PES); exercise; medication misuse; pain burden | NA | Increase | Increase | NA |
| Weymann et al [18] | heiQ; patient knowledge; decisional conflict; preparation for decision making | NA | No effect | NA | NA |

[a]BPI: Brief Pain Inventory; CPCI-42: Chronic Pain Coping Inventory; CSQ: Coping Strategies Questionnaire; DASS: Depression Anxiety Stress Scale; FABQ: Fear-Avoidance Beliefs Questionnaire; HADS: Hamilton Anxiety and Depression Scale; heiQ: Health Education Impact Questionnaire; KPD-38: Clinical Psychological Diagnostic System; MOS: Medical Outcomes Study; MPI: Multidimensional Pain Inventory; NMRS: Negative Mood Regulation Scale; NRS: Numeric Rating Scale; PAIRS: Pain and Impairment Relationship Scale; PCS: Pain Catastrophizing Scale; PES: Psychological Empowerment Scale; PGIC: Patients' Global Impression of Change Scale; PSEQ: Pain Self-efficacy Questionnaire; QOLI: Quality of Life Inventory; RMQ: Roland-Morris Disability Questionnaire; SES: Pain Self-efficacy Scale; SOPA: Survey of Pain Attitudes; VNS: Visual Numeric Scale.

[b]In intervention group. NA: not available.

The studies were presented in two subsections: studies using online cognitive behavioral therapy (CBT) and studies using Web-based approaches to improve knowledge (with an interactive component to provide coping support).

The following trials were registered: Burhman et al [13], Krein et al [16], and Riva et al [17].

### Studies Using Online Cognitive Behavioral Therapy

Psychological factors, such as depressed mood, negative beliefs, and somatization, have been shown to affect chronicity of pain and disability related to the pain [19,20]. This review identified four RCTs published between 2004 and 2012 that examined the effectiveness of Internet-based CBT as part of the treatment strategy for chronic back pain. The number of participants randomized in each study varied between 54 and 228. The majority of participants in all four studies were women (62.5%-83%).

Burhman et al [11] used Internet-based CBT in conjunction with telephone support to treat chronic back pain. The study reported that 95 participants would be required for a power of 80%; however, due to lower enrollment the study remained underpowered. The primary outcome measure was catastrophization, defined as the experience of irrationally thinking that something is far worse than it actually may be. This was measured as a subscale of the Coping Strategies Questionnaire (CSQ). The CSQ consisted of measurements of the following parameters: diverting attention, reinterpret pain sensations, coping self-statements, ignore pain sensations, praying or hoping, catastrophizing, increase activity level, control over pain, and ability to decrease pain. Patients were randomized to Web-based pain management or a waitlist control. The intervention group received access to weekly online CBT modules, guidance with physical activity and stretching exercises, and coping strategies over the course of 12 weeks. The intervention group also received weekly telephone calls that included discussion about participant goals, relaxation training advice, exercise guidance, and discussion on coping strategies. These calls occurred during the same period as the online intervention. The treatment group showed lower tendency to catastrophize and also reported better control over pain at 8 weeks. Due to significant follow-up through telephone calls, it is unclear how much of the treatment effect can be attributed



XSL•FO
RenderX

to the online modality of treatment as opposed to the telephone-based support.

Burhman et al [13] performed a similar study to the one described previously, but without ongoing telephone support as part of the treatment plan. In all, 54 patients were randomized; it was reported that the study was underpowered to detect differences with conventional levels of confidence. The treatment group in this study also showed improved scores on the catastrophization subscale. Although this study did not include telephone support as part of the intervention, ongoing email support was provided to participants. Therefore, it is not entirely clear how much of the treatment effect can be attributed to the online CBT modules as opposed to the effect of email support. Along with reduction in catastrophization, the participants also reported improved control over pain with the intervention. Because the current paradigm of chronic pain management stresses the importance of maintenance or improvement in patient function, this may be seen as an important initial step toward achieving better self-efficacy and an improvement in the patient's ability to understand and manage their own pain. However, further research would be useful to clarify whether this may indeed translate into improvements in pain and disability scores.

Carpenter et al [14] also studied an online self-help CBT intervention. The study included 141 participants who had back pain for more than 6 months and were older than 21 years of age. Over the course of three weeks, the treatment group used an online wellness workbook that included elements of CBT; the results reported an improved ability to self-manage pain. The treatment group reported decreased pain catastrophizing and a more positive outlook toward their disability. After week 3, the treatment group reported an improvement in their perceived ability to cope with their pain. Conversely, the participants in the control group were less confident about their ability to manage pain and were more likely to believe that they should avoid exercise. The study then allowed both groups to access the online workbook after the 3-week period and repeated their assessments for all the participants at 6 weeks. It was reported that the differences in the two groups were no longer apparent at 6 weeks, suggesting that access to the workbook successfully affected participants' pain-related beliefs.

Chaiuzzi et al [12] compared an intervention group with access to a website (painACTION for back pain) designed on CBT self-management principles with a control group of participants provided with a back pain help book. Participants were recruited online and through a specialty pain clinic. The sample size was 228; sample size and power calculations were not reported. The intervention group received access to the CBT website and a weekly chat moderated by a therapist. Posttreatment follow-up at 3 and 6 months was performed. Overall, the intervention group reported reduced stress and improved coping, but pain and physical functioning were not affected significantly. However, in a subgroup of patients recruited online, pain levels did appear to be improved with the intervention compared to the control group.

In summary, four small RCTs reporting the effects of Web-based CBT for chronic back pain have been identified. All studies found reduced catastrophization in patients receiving online CBT.

Each of the trials used different measures to report pain levels. These measures included the Pain and Impairment Relationship Scale (PAIRS), Pain Self-efficacy Scale (SES), a pain diary, and self-reported pain levels for least, average, and worst pains. Of the studies that examined CBT, only Carpenter et al [14] used CBT as an adjunct to opioid therapy. None of the studies reported significant differences in pain severity.

### Studies Using Web-Based Approaches and an Interactive Component

Web-based interventions with interactive features are being increasingly studied for their potential role in the management of chronic diseases. The results from a recent review indicated Web-based interactive interventions for patients with a variety of chronic conditions may have a positive impact on patient empowerment and may facilitate enhanced physical activity [21].

The studies discussed in this section target knowledge about chronic low back pain by providing online resources and also provide support for coping through Web-based interactive features.

Lorig et al [10] performed an RCT to examine the impact of participation in email discussion groups; the outcomes of interest were health status and health care utilization. Study duration was 1 year and 580 participants were randomly assigned to treatment and control groups. The intervention group was enrolled in an email discussion group where various aspects of back pain were discussed with input from content experts. The content experts included a physician, physical therapist, and psychologist. This study used moderated email discussion; however, the topics for email discussion were mostly driven by participants and no specific predesigned content was provided to the participants. Further, the intervention group also received a back pain help book and a videotape modeling active living with back pain. No particular physical activity routine or exercise was suggested; rather, the email discussion answered general questions raised by the participants. The control group did not receive any specific back pain treatment or advice. The study included a 6-month and 1-year follow-up. At 1 year, improvements in pain, disability, role function, and health distress were reported with the intervention. The study was powered to detect these differences with a significance of $P<.05$. Health care utilization was reduced in the treatment group, but not to a statistically significant degree. The number of physician visits were decreased in the treatment group. Further, the mean number of hospital days (back-related days of hospitalization) were reduced by 0.25 days for the intervention group as compared to an increase of 0.04 days for the control group. Self-care orientation was improved with treatment. The study also reported that older age was associated with greater disability. The authors indicate and recognize that there are multiple factors affecting pain levels and health care utilization due to chronic back pain; consequently, it is unclear how the results can be attributed to the various parts of the intervention (discussion group, back pain help book, and videotape).



XSL•FO
RenderX

Moessner et al [15] studied aftercare intervention for patients who had already received multidisciplinary therapy for back pain. The study randomized 75 patients; low power was reported due to small sample size. Participants received an Internet-based aftercare intervention lasting 15 weeks or treatment as usual. The aftercare program included an individualized online self-monitoring module, where participants answered questions about their compliance with appropriate health behaviors. Also, the aftercare consisted of a 90-minute weekly text-based chat for a period of 15 weeks. The chat was moderated by an experienced group therapist; session topics were decided by the therapist. A physician or physiotherapist were not included as moderators for this chat. The results reported improvements in disability with the intervention. No significant difference in depression or anxiety was reported. Despite the positive results, there is an important caveat: a significant amount of data was lost because only 34 of 75 patients completed all three assessments. Moreover, the authors did not report the components of the multidisciplinary rehabilitation; as such, there is no way of knowing whether—and the degree to which—results of their Internet-based intervention were affected by components of rehabilitation. The authors theorized that patient beliefs about chronic pain may have impacted follow-up.

Krein et al [16] conducted a study that focused on improving the activity level of the participants by providing them with pedometers that gave online feedback regarding their daily activity. The pedometer feedback was used in combination with an e-community social support group. The researchers randomized 229 patients into a control group and a treatment group. The study was designed to detect a clinically meaningful difference (0.4 standard deviation or 2-point difference) in Roland-Morris Disability Questionnaire score and sought to enroll 130 participants in each group to account for potential of 25% attrition; high rates of participant follow-up were achieved for this study allowing for detection of differences in primary outcome. The treatment group received pedometers along with access to online feedback including the number of steps and individual goals to promote improvement. The treatment group was also provided with access to an online social support group. In contrast, the control group received pedometers but did not receive online feedback or social support. Assessments were performed at 6- and 12-month time points. Most participants were males. Significant improvement was reported for the treatment group compared to the control group for back pain disability at 6 months, but the difference was no longer statistically significant at 12 months. No difference was reported between the groups in terms of Fear-Avoidance Beliefs Questionnaire Physical Activity subscale. Physical activity measured by step counts was increased in the intervention group at 6 months; however, this difference was less marked when measured at 12 months. Exercise self-efficacy scores were similar between the two groups at 12 months.

Riva et al [17] randomized 51 patients into two groups. It was reported that the study was designed to achieve power of 80% with 95% confidence, and sufficient numbers were recruited for this purpose. The intervention group received access to a self-management website with interactive components including quizzes, virtual gym, an action plan, and additional online resources. The control group only received access to static features and information on the website. Four- and 8-week assessments were performed. Outcome measures included empowerment, medication misuse, physical exercise, and pain burden. The intervention group was reported to have improved patient empowerment and reduced medication misuse. Pain burden decreased, but to equal measures in both the control and intervention groups. Because pain levels decreased in both groups, it appears that the interactive features available to the intervention group did not make a significant difference to their pain levels. However, participant empowerment was reported to be significantly improved in the intervention group. It appeared that interactivity and feedback through the Internet may improve a sense of control or empowerment in chronic back pain patients.

Weymann et al [18] included participants with chronic low back pain and type 2 diabetes in their study. A total of 561 participants were randomized, of which 382 were enrolled with chronic back pain. The intervention was a tailored interactive health communication app, which provided support to participants with regards to their knowledge and attitudes about their condition. The coping style of participants was assessed prior to intervention; participants in the intervention group were offered tailored content based on their coping style. For chronic low back pain participants, information was based on recent guidelines and Cochrane reviews. Primary outcomes were patient knowledge and patient empowerment.

The study aimed to detect differences with conventional levels of confidence and 80% power; however, due to attrition, only 202 of 382 chronic low back pain participants performed the 3-month follow-up. No significant differences were detected in outcomes with the intention-to-treat analysis.

Two of the studies [10,15] reported reduction in disability. Krein et al [16] reported reduction in disability at 6 months, which was not sustained in further assessments. Lorig et al [10] reported statistically significant reduction in pain, whereas Moessner et al [15] reported improvement with the pain subscale of the 36-item Short-Form Health Survey (SF-36), but not with Numeric Rating Scale (NRS).

Four of the studies did measure empowerment/self-efficacy and mixed results were reported. Empowerment was reported to be improved in one of the studies [17] and another study reported improved self-efficacy [10]. However, no difference was reported in other studies [16,18].

A variety of diverse outcome measures have been used in the studies. Lorig et al [10] did measure health care utilization, which was reported to be decreased in their treatment group. In the context of chronic low back pain, this outcome measure has not been extensively studied in RCTs since this trial; it would be prudent for future researchers to include cost or health care utilization as an outcome measure. Further, based on the study design, it may not be possible to ascertain the individual contribution of each part of the intervention to the reported outcomes. Also, the discussion is not supportive of any particular physical activity intervention and no individual medical advice was provided to participants. This suggests that participant



XSL•FO
RenderX

self-efficacy may independently affect outcomes in this population.

## Discussion

Nine unique RCTs were identified addressing the impact of Web-based interventions on chronic low back pain. The major categories of interventions included online CBT and to improve knowledge with an interactive component to provide coping support. The trials identified had small sample sizes and many of them were not blinded. In terms of power calculations, three of the trials reported being underpowered. There is considerable concern with external validity for these study results. The demographics of the population included for the different studies were heterogenous. The delivery, format, and timeline of the interventions were also heterogenous. Most studies only reported posttreatment data and there is a lack of long-term follow-up. In the studies that do report longer-term data, the treatment effects seem to taper off with time [16].

Many of the studies excluded patients receiving disability payments, a significant part of the population that experiences chronic back pain. As such, the absence of research on this subpopulation is a major gap that should be addressed in future studies. The effect of Web-based interventions on health care utilization was reported by only one study [10] and indicated a trend toward reduced physician visits for back pain. This is an important outcome measure that would be useful to include in future studies to better understand effects of online interventions on health care access, system burden, and resources. CBT has been linked to improved outcomes in many chronic conditions and this review indicates that CBT has been effective for chronic pain. However, specific mechanisms through which the CBT treatment has its effect are not entirely clear and more research on this process is necessary [22]. Several studies have reported decreases in catastrophization and/or improvement in self-efficacy and this may lead to improvements in health-related behaviors or follow-up and adherence with appropriate treatments. Also, patient characteristics that make them more likely to respond to CBT have not been adequately studied [22].

Four RCTs reporting the effects of Web-based CBT for chronic back pain were identified for this review. Three of the studies report reduced catastrophization in patients receiving online CBT. In previous studies on chronic pain, catastrophization has been linked to increased severity of pain, poor treatment outcomes, and increased disability [1]. However, one limitation to the online CBT studies is that the majority of participants were women. Researchers have previously noted the women seek health care more often for pain compared to men [23]. Further, the incidence of low back pain appears to be higher among females and those aged between 40 and 80 years [24]. However, it is important to have studies with more male participants—or a mix of demographics—to improve the applicability and generalizability of the results. Also, the format and the dose of CBT provided in different trials are variable. Therefore, it is difficult to draw conclusions regarding the optimal frequency, duration, and format of CBT that may be required to improve outcomes in chronic back pain.

Further, there are various limitations to the studies using online CBT. All studies randomized small numbers of patients at single centers. Some of the studies are not adequately powered. One study was waitlist controlled, which can be problematic because this can make the results of the treatment effect appear more significant than it actually is. Intention-to-treat analysis was not conducted in two of the studies; therefore, participants with suboptimal compliance are excluded from parts of the analysis. Also, the form and type of delivery of supports in addition to online intervention were variable. For example, Burhman et al [11] made significant use of telephone support and provided consistent advice regarding physical activity, whereas Carpenter et al [14] focused solely on behavior and cognitive exercises. Therefore, effects found in Burhman et al [11] may be attributed to multiple interventions rather than CBT alone.

Another limitation is that most of the participants in the studies were females, which may affect the generalizability of the results. Some of the studies excluded patients with comorbidities such as heart disease; this may affect how results can be interpreted because many patients seen in practice with low back pain have significant comorbidities, which may also limit generalizability. In general, the samples in the studies may have been so carefully selected that their external validity is questionable.

Further, none of the studies included groups receiving face-to-face CBT as controls; therefore, it is not possible to estimate the efficacy of online CBT in comparison with the traditional approach. Overall, the research indicates that online CBT may be effective in reducing catastrophization and improving patient attitudes toward back pain, particularly when supported with telephone or email follow-up. However, additional RCTs with larger and more diverse samples are required to further investigate whether this intervention can be effective in reducing pain, disability, and health care costs. Furthermore, studies must be conducted to consider independent effects from total effects for each aspect of treatment.

Five RCTs reporting effects of Web-based approaches to improve knowledge and coping support. Three of these studies reported a reduction in disability [10,15,16], although in one study this benefit was not sustained on assessment at 1 year [16]. Two of the studies also appear to show improvement in pain levels [10,15]. Empowerment was reported to be improved in one of the studies [17] and another study reported improved self-efficacy [10].

These studies have a number of limitations. One omission is that Riva et al [17] did not include information on whether their samples were using medications to reduce their back pain over the course of their studies; however, it may be assumed that many patients with chronic back pain will access or use medications. In terms of patient empowerment and patient self-efficacy, mixed results were reported [16-18]. These two constructs are conceptually related; as such, we would expect to find similar effects of the online interactive intervention across both samples. Further, the studies have small sample sizes and three of the studies are underpowered. A majority of the participants in the study by Krein et al [16] were male, which can affect the generalizability of the results. Many of the studies



XSL•FO
RenderX

have short duration and/or follow-up. Therefore, it may be difficult to draw conclusions regarding the long-term impact of the interventions, especially with regards to function and disability.

## Conclusions

Although research on many of the Web-based interventions for back pain reviewed here had mixed results or do not appear to have high external validity, we did find evidence that that there are likely some benefits to online CBT for reduced catastrophization. As such, online interventions may be a useful solution to overcome current limitations of traditional face-to-face CBT because, for example, access to professionals that are able to deliver high-quality CBT remains limited. Second, many patients may not be able to afford access to such professionals or counseling. Third, physical access may also be limited due to the nature of pain, patient comorbidities, or other social factors, and large geographical distances may preclude eligible patients from accessing specialized rehabilitation or chronic pain centers. Fourth, in some cases, there may be a stigma associated with the use of a therapist or counselor. Therefore, online access to CBT may help to alleviate some of the barriers to access and provide patients a convenient alternative to face-to-face visits. Future studies using CBT as an intervention should consider including appropriate numbers of male participants to improve the generalizability of the results.

Further, empowerment/control did show improvement in six of the studies. Three of these used CBT, whereas three of the studies used other forms of Web-based support, such as email/chat or other interactive features. It appears that forms of social support other than formalized counseling or CBT may have some positive effect on the patient's ability to manage and cope with their chronic condition.

Disability was only assessed in five of the studies and mixed results were reported. Further research in this area with studies having longer follow-up should be a priority. One study reporting health care utilization reported positive effects with the intervention. It would be important for future studies to assess this further because it is important to focus resources on interventions that can reduce use of health care resources. Further research that includes these outcomes could provide insight into future planning for the health care system and implications for clinical practice.

## Acknowledgments

The authors wish to acknowledge support from the Department of Family Medicine at the University of Calgary.

## Conflicts of Interest

None declared.

## References

1. Andersson GB. Epidemiological features of chronic low-back pain. Lancet 1999 Aug 14;354(9178):581-585. [doi: 10.1016/S0140-6736(99)01312-4] [Medline: 10470716]
2. Strine TW, Hootman JM. US national prevalence and correlates of low back and neck pain among adults. Arthritis Rheum 2007 May 15;57(4):656-665 [FREE Full text] [doi: 10.1002/art.22684] [Medline: 17471542]
3. Edwards RR, Bingham CO, Bathon J, Haythornthwaite JA. Catastrophizing and pain in arthritis, fibromyalgia, and other rheumatic diseases. Arthritis Rheum 2006 Apr 15;55(2):325-332 [FREE Full text] [doi: 10.1002/art.21865] [Medline: 16583384]
4. Baldwin ML. Reducing the costs of work-related musculoskeletal disorders: targeting strategies to chronic disability cases. J Electromyogr Kinesiol 2004 Feb;14(1):33-41. [doi: 10.1016/j.jelekin.2003.09.013] [Medline: 14759748]
5. NIH National Institute of Neurological Disorders and Stroke. URL: http://www.ninds.nih.gov/ [accessed 2016-05-09] [WebCite Cache ID 6hNeQeucW]
6. Dooley D, Fielding J, Levi L. Health and unemployment. Annu Rev Public Health 1996;17:449-465. [doi: 10.1146/annurev.pu.17.050196.002313] [Medline: 8724235]
7. Williams RM, Westmorland M. Perspectives on workplace disability management: a review of the literature. Work 2002;19(1):87-93. [Medline: 12454354]
8. Guzmán J, Esmail R, Karjalainen K, Malmivaara A, Irvin E, Bombardier C. Multidisciplinary rehabilitation for chronic low back pain: systematic review. BMJ 2001 Jun 23;322(7301):1511-1516 [FREE Full text] [Medline: 11420271]
9. van Middelkoop M, Rubinstein SM, Kuijpers T, Verhagen AP, Ostelo R, Koes BW, et al. A systematic review on the effectiveness of physical and rehabilitation interventions for chronic non-specific low back pain. Eur Spine J 2011 Jan;20(1):19-39 [FREE Full text] [doi: 10.1007/s00586-010-1518-3] [Medline: 20640863]
10. Lorig KR, Laurent DD, Deyo RA, Marnell ME, Minor MA, Ritter PL. Can a back pain e-mail discussion group improve health status and lower health care costs?: a randomized study. Arch Intern Med 2002 Apr 8;162(7):792-796. [Medline: 11926853]
11. Buhrman M, Fältenhag S, Ström L, Andersson G. Controlled trial of Internet-based treatment with telephone support for chronic back pain. Pain 2004 Oct;111(3):368-377. [doi: 10.1016/j.pain.2004.07.021] [Medline: 15363881]



XSL•FO
RenderX

12. Chiauzzi E, Pujol LA, Wood M, Bond K, Black R, Yiu E, et al. painACTION-back pain: a self-management website for people with chronic back pain. Pain Med 2010 Jul;11(7):1044-1058 [FREE Full text] [doi: 10.1111/j.1526-4637.2010.00879.x] [Medline: 20545873]
13. Buhrman M, Nilsson-Ihrfeldt E, Jannert M, Ström L, Andersson G. Guided internet-based cognitive behavioural treatment for chronic back pain reduces pain catastrophizing: a randomized controlled trial. J Rehabil Med 2011 May;43(6):500-505 [FREE Full text] [doi: 10.2340/16501977-0805] [Medline: 21533329]
14. Carpenter KM, Stoner SA, Mundt JM, Stoelb B. An online self-help CBT intervention for chronic lower back pain. Clin J Pain 2012 Jan;28(1):14-22 [FREE Full text] [doi: 10.1097/AJP.0b013e31822363db] [Medline: 21681084]
15. Moessner M, Schiltenwolf M, Neubauer E. Internet-based aftercare for patients with back pain-a pilot study. Telemed J E Health 2012;18(6):413-419. [doi: 10.1089/tmj.2011.0221] [Medline: 22676398]
16. Krein SL, Kadri R, Hughes M, Kerr EA, Piette JD, Holleman R, et al. Pedometer-based internet-mediated intervention for adults with chronic low back pain: randomized controlled trial. J Med Internet Res 2013;15(8):e181 [FREE Full text] [doi: 10.2196/jmir.2605] [Medline: 23969029]
17. Riva S, Camerini A, Allam A, Schulz PJ. Interactive sections of an Internet-based intervention increase empowerment of chronic back pain patients: randomized controlled trial. J Med Internet Res 2014;16(8):e180 [FREE Full text] [doi: 10.2196/jmir.3474] [Medline: 25119374]
18. Weymann N, Dirmaier J, von Wolff A, Kriston L, Härter M. Effectiveness of a Web-based tailored interactive health communication application for patients with type 2 diabetes or chronic low back pain: randomized controlled trial. J Med Internet Res 2015;17(3):e53 [FREE Full text] [doi: 10.2196/jmir.3904] [Medline: 25736340]
19. Pincus T, Burton AK, Vogel S, Field AP. A systematic review of psychological factors as predictors of chronicity/disability in prospective cohorts of low back pain. Spine (Phila Pa 1976) 2002 Mar 1;27(5):E109-E120. [Medline: 11880847]
20. Vargas-Prada S, Martínez JM, Coggon D, Delclos G, Benavides FG, Serra C. Health beliefs, low mood, and somatizing tendency: contribution to incidence and persistence of musculoskeletal pain with and without reported disability. Scand J Work Environ Health 2013 Nov;39(6):589-598 [FREE Full text] [doi: 10.5271/sjweh.3377] [Medline: 23955508]
21. Kuijpers W, Groen WG, Aaronson NK, van Harten WH. A systematic review of web-based interventions for patient empowerment and physical activity in chronic diseases: relevance for cancer survivors. J Med Internet Res 2013;15(2):e37 [FREE Full text] [doi: 10.2196/jmir.2281] [Medline: 23425685]
22. Ehde D, Dillworth T, Turner JA. Cognitive-behavioral therapy for individuals with chronic pain: efficacy, innovations, and directions for research. Am Psychol 2014;69(2):153-166. [doi: 10.1037/a0035747] [Medline: 24547801]
23. Unruh AM. Gender variations in clinical pain experience. Pain 1996;65(2-3):123-167. [Medline: 8826503]
24. Hoy D, Bain C, Williams G, March L, Brooks P, Blyth F, et al. A systematic review of the global prevalence of low back pain. Arthritis Rheum 2012 Jun;64(6):2028-2037 [FREE Full text] [doi: 10.1002/art.34347] [Medline: 22231424]


## Abbreviations

**BPI:** Brief Pain Inventory
**CBT:** cognitive behavioral therapy
**CPCI-42:** Chronic Pain Coping Inventory
**CSQ:** Coping Strategies Questionnaire
**DASS:** Depression Anxiety Stress Scale
**FABQ:** Fear-Avoidance Beliefs Questionnaire
**HADS:** Hamilton Anxiety and Depression Scale
**heiQ:** Health Education Impact Questionnaire
**MOS:** Medical Outcomes Study
**MDT:** multidisciplinary team
**MPI:** Multidimensional Pain Inventory
**NMRS:** Negative Mood Regulation Scale
**NRS:** Numeric Rating Scale
**PAIRS:** Pain and Impairment Relationship Scale
**PCS:** Pain Catastrophizing Scale
**PES:** Psychological Empowerment Scale
**PGIC:** Patients' Global Impression of Change Scale
**PSEQ:** Pain Self-efficacy Questionnaire
**QOLI:** Quality of Life Inventory
**RCT:** randomized controlled trial
**RMQ:** Roland-Morris Disability Questionnaire
**SES:** Self-Efficacy Scale
**SOPA:** Survey of Pain Attitudes
**VNS:** Visual Numeric Scale



XSL•FO
RenderX

*Edited by G Eysenbach; submitted 12.07.15; peer-reviewed by A Allam, W Kuijpers, S Riva, P Schulz; comments to author 03.11.15; revised version received 24.01.16; accepted 20.02.16; published 26.07.16.*

*Please cite as:*
*Garg S, Garg D, Turin TC, Chowdhury MFU*
*Web-Based Interventions for Chronic Back Pain: A Systematic Review*
*J Med Internet Res 2016;18(7):e139*
*URL: http://www.jmir.org/2016/7/e139/*
*doi:10.2196/jmir.4932*
*PMID:27460413*

©Shashank Garg, Divya Garg, Tanvir C Turin, M Faruq U Chowdhury. Originally published in the Journal of Medical Internet Research (http://www.jmir.org), 26.07.2016. This is an open-access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/2.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on http://www.jmir.org/, as well as this copyright and license information must be included.



XSL•FO
RenderX

Original Paper

# A Mobile Ecological Momentary Assessment Tool (devilSPARC) for Nutrition and Physical Activity Behaviors in College Students: A Validation Study

Meg Bruening[1], RD, MPH, PhD; Irene van Woerden[1], MSc; Michael Todd[1], PhD; Stephanie Brennhofer[1], MSc, RDN; Melissa N Laska[2], RD, PhD; Genevieve Dunton[3], MPH, PhD

[1]Arizona State University, Phoenix, AZ, United States

[2]University of Minnesota, Minneapolis, MN, United States

[3]University of Southern California, Los Angeles, CA, United States

**Corresponding Author:**
Meg Bruening, RD, MPH, PhD
Arizona State University
550 N 5th Street
Phoenix, AZ, 85004
United States
Phone: 1 602 827 2266
Fax: 1 602 827 2254
Email: meg.bruening@asu.edu

## Abstract

**Background:** The majority of nutrition and physical activity assessments methods commonly used in scientific research are subject to recall and social desirability biases, which result in over- or under-reporting of behaviors. Real-time mobile-based ecological momentary assessments (mEMAs) may result in decreased measurement biases and minimize participant burden.

**Objective:** The aim was to examine the validity of a mEMA methodology to assess dietary and physical activity levels compared to 24-hour dietary recalls and accelerometers.

**Methods:** This study was a pilot test of the SPARC (Social impact of Physical Activity and nutRition in College) study, which aimed to determine the mechanism by which friendship networks impact weight-related behaviors among young people. An mEMA app, devilSPARC, was developed to assess weight-related behaviors in real time. A diverse sample of 109 freshmen and community mentors attending a large southwestern university downloaded the devilSPARC mEMA app onto their personal mobile phones. Participants were prompted randomly eight times per day over the course of 4 days to complete mEMAs. During the same 4-day period, participants completed up to three 24-hour dietary recalls and/or 4 days of accelerometry. Self-reported mEMA responses were compared to 24-hour dietary recalls and accelerometry measures using comparison statistics, such as match rate, sensitivity and specificity, and mixed model odds ratios, adjusted for within-person correlation among repeated measurements.

**Results:** At the day level, total dietary intake data reported through the mEMA app reflected eating choices also captured by the 24-hour recall. Entrées had the lowest match rate, and fruits and vegetables had the highest match rate. Widening the window of aggregation of 24-hour dietary recall data on either side of the mEMA response resulted in increased specificity and decreased sensitivity. For physical activity behaviors, levels of activity reported through mEMA differed for sedentary versus non-sedentary activity at the day level as measured by accelerometers.

**Conclusions:** The devilSPARC mEMA app is valid for assessing eating behaviors and the presence of sedentary activity at the day level. This mEMA may be useful in studies examining real-time weight-related behaviors.

***(J Med Internet Res 2016;18(7):e209)*** doi:10.2196/jmir.5969

**KEYWORDS**

validation study; ecological momentary assessment; nutritional status; physical activity; sedentary activity; emerging adults



XSL•FO
RenderX

## Introduction

The majority of nutrition and physical activity (PA) assessments are subject to recall and social desirability biases, which can result in over- or under-reporting of behaviors [1,2]. For example, studies of dietary intake in adolescents and young adults have shown that people generally overestimate or underestimate their own consumption [3]. When self-reporting PA behaviors, young people tend to overestimate the time spent in, and the intensity of, PA efforts [4,5]. As such, many nutrition and PA measures suffer from low validity [6], resulting in limited interpretability of findings. There is a need to understand the nutrition and PA behaviors of young people through reliable and valid measurement tools that do not impose a high level of burden on participants or high costs to researchers.

Ecological momentary assessments (EMAs) limit measurement biases associated with self-reported recall data. As described by Shiffman et al [7] and Stone et al [8], EMAs involve sampling strategies that assess phenomena in the moment they occur in the participant's natural environment, and they have at least three major advantages over traditional measurement tools for diet and PA: (1) avoidance of recall bias by collecting data in real time or near real time; (2) maximizing ecological validity by assessing behaviors in the environments where they occur; and (3) fine-grained temporal resolution, enabling analysis of behavior as it unfolds over time. A review of studies comparing EMA with traditional long-term recall-based methods points to EMA being better able to generate more valid results when researchers are interested in understanding a person's experience as it occurs rather than their retrospective impressions of the experience [7]. In the context of eating and PA behavior assessment, the use of EMA could also lead to decreased participant burden, potentially yielding higher rates of compliance and lower rates of missing data.

Emerging technologies and changes in how people use technologies have created opportunities to assess behaviors as the behaviors occur. With increasing use and ownership of mobile phones, mobile technologies are valuable assets in behavioral health research [2,9,10]. Several studies have shown the usefulness and effectiveness of mobile technology-based EMAs (mEMAs), including mobile phone apps, texting, and personal digital assistants (PDAs) [10-15]; mEMAs are almost exclusively used in EMA research these days (as opposed to paper-and-pen and desktop computer EMAs), particularly in research with young people [15].

College students are an understudied population in regards to weight and weight-related behaviors and management [16]. The National Institutes of Health (NIH) has encouraged technology-driven weight management interventions for young adults due to the deficit of efforts for this critical population in transition [17,18]. mEMAs, particularly those using mobile phones, may be particularly useful for studying the behavior of today's young adults, as these youth tend to make frequent and extensive use of mobile phones, owing in part to being the first generation to grow up completely with mobile technologies [19]. Because of their development stage, stressors, and ever-changing priorities, college students can be a difficult population to study over time [20], and finding ways to maximize compliance is critical. The use of mEMA in young adult research is limited, especially when focusing on nutrition and PA. Studies using mEMA methods in young people have tended to focus on substance use and other harmful behaviors (eg, tobacco use, marijuana use, binge eating) [11]. A few studies have validated mEMA for PA assessment in elementary and adolescent age groups [2,12]; however, to our knowledge, no study has validated mEMA for PA assessment in older adolescent/young adult (aged 16-21 years) populations or for dietary assessment in any population. In this study, we sought to examine the validity of a mEMA app, *devilSPARC*, in assessing dietary and PA behaviors with a college student sample.

## Methods

### Study Design

The SPARC (Social impact of Physical Activity and nutRition in College) study was a large-scale NIH-funded study that aimed to determine the mechanisms by which friendship networks and interpersonal connections impact weight and weight-related outcomes. In the formative phase of the study, participants answered EMA prompts asking about their current nutrition and PA behaviors using the mEMA app, devilSPARC, as well as validated measures of diet [20] and/or accelerometry across a 4-day period during the 2014-2015 academic year. Participants provided written consent prior to enrollment and were offered incentives of up to US $80 for their completion of the pilot study. All study protocols were approved by Arizona State University's Institutional Review Board.

### Participants

College freshmen and assigned community mentors (resident assistants) at Arizona State University from two residence halls were recruited for participation. Inclusion criteria were (1) enrollment at Arizona State University and (2) living in target residence halls. For those participants who were interested in participating but did not own an Android or iOS mobile phone, a Motorola Moto G was loaned to them for use during the duration of the study. The resulting sample was 109 participants: 68 participants who provided mEMA and 24-hour dietary recalls only, 17 students who provided mEMA PA reports and accelerometry assessments only, and 24 participants who completed both protocols (92 dietary recalls, 41 accelerometry assessments).

### Measures

#### *The devilSPARC Mobile Ecological Momentary Assessment App*

The mEMA software was designed specifically for this study and implemented on Android- or iOS-compatible mobile phones. For each 4-day period of data collection (Wednesday-Saturday), participants received a total of 32 short message service (SMS) text message prompts to complete mEMA surveys via the devilSPARC app. Each day, participants received seven "real-time" prompts per day (n=28 total) and one retrospective prompt per day (n=4). Real-time prompts asked participants



XSL•FO
RenderX

what they were doing in the moment before they received the prompt and retrospective prompts asked participants to recall what they did in the past 3 hours. A random, interval-contingent schedule was used for the mEMA prompts. Twice during each of the four established time periods per day (9 am-12 pm, 12 pm-3 pm, 3 pm-7 pm, and 7 pm-10 pm) the system prompted participants to complete a brief survey. In order to ensure the momentary nature of the mEMA, participants were allotted 35 minutes to respond to the prompt by completing a 1-minute survey, with the survey being available for 5 minutes prior to, and 30 minutes after, the text message prompt. Outside of these times, the mEMA surveys were not available to complete on the app. On average, the latency time (time between the sending of the SMS prompt to the completion of the survey on the mEMA app) was 7.25 minutes for participants completing the dietary validation and 6.90 minutes for participants completing the PA validation. Trained research assistants downloaded the devilSPARC mEMA app to each participant's mobile phone and provided demonstrations on how to use the devilSPARC app.

All SMS text prompts were sent directly to participants' mobile phones using Web service application programming interfaces (APIs) provided by Twilio, a cloud communications company. Through Twilio's API, a series of six local long codes (10-digit phone numbers) were used to send text messages to participants. Each long code was randomly assigned to a participant based on his or her participant identification number. Any failed text messages were sent two more times, for a total of three attempts. The text message included motivational text with an embedded link that would open a survey on the app. Data were transferred instantaneously to the study host server every time the user's phone contacted the central server (eg, on submission of a survey or opening of the home screen).

Figure 1 includes screenshots of the real-time mEMA items, which included the assessment of eating, drinking, PA and sedentary behaviors, and activities. The sequence of items measured varied based on a participant's response to the first question, "What were you doing right before you got this text?" Participants could select all of the following that applied: eating, drinking, being physically active, or none of the above. If participants selected eating, then they were asked to identify food groups they were eating: (1) cookies/sweetened baked goods/candy/frozen desserts (sweets); (2) salty snacks/fried side dishes (salty foods); (3) fruits and vegetables; (4) entrées, (5) breads, cereals, and grains (breads/grains); and (6) other. These food groups were identified in previous research as types of food frequently consumed by college students [21]. If participants selected PA, then an adapted version of the Godin-Shepard measure of self-reported PA [22] was shown, and participants were instructed, "Select the activity that most closely matches what you are doing: strenuous exercise (heart beats rapidly), moderate exercise (not exhausting), and mild exercise (little effort)." If a participant did not select that they were being physically active, then it was assumed that they were involved in sedentary activity. Participants were then asked to respond to the following item: "Select any activity that most closely matches what you are doing (not including responding to this assessment)." Response options included sleeping, browsing the Internet, using social media, watching TV or a movie, playing video games, texting/snapchatting, attending class/doing homework/studying/reading, working, hanging out, and other (specify).

### *Dietary Recall*

The online version of the Automated Self-Administered 24-hour (ASA24) dietary recall system, a validated measure of self-reported dietary intake, was used to assess participants' food and beverage intake over the previous 24 hours [23]. The ASA24 utilizes the US Department of Agriculture's (USDA) Automated Multiple Pass Method and measures intake by using the USDA's Food and Nutrient Database for Dietary Studies. Participants were asked to complete 3 days of dietary recall (two weekdays and one weekend day). If participants reported at least one full day of biologically plausible data (ie, daily caloric intake between 500-5000 kilocalories [24-27]), their data were used in the analytic sample. Each food item reported in the ASA24 was coded to match the food groups in the mEMA: (1) sweets, (2) salty snacks/fried side dishes, (3) fruits and vegetables, (4) entrées (eg, pizza, sandwiches, lasagnas, chicken), (5) breads/grains, and (6) other.

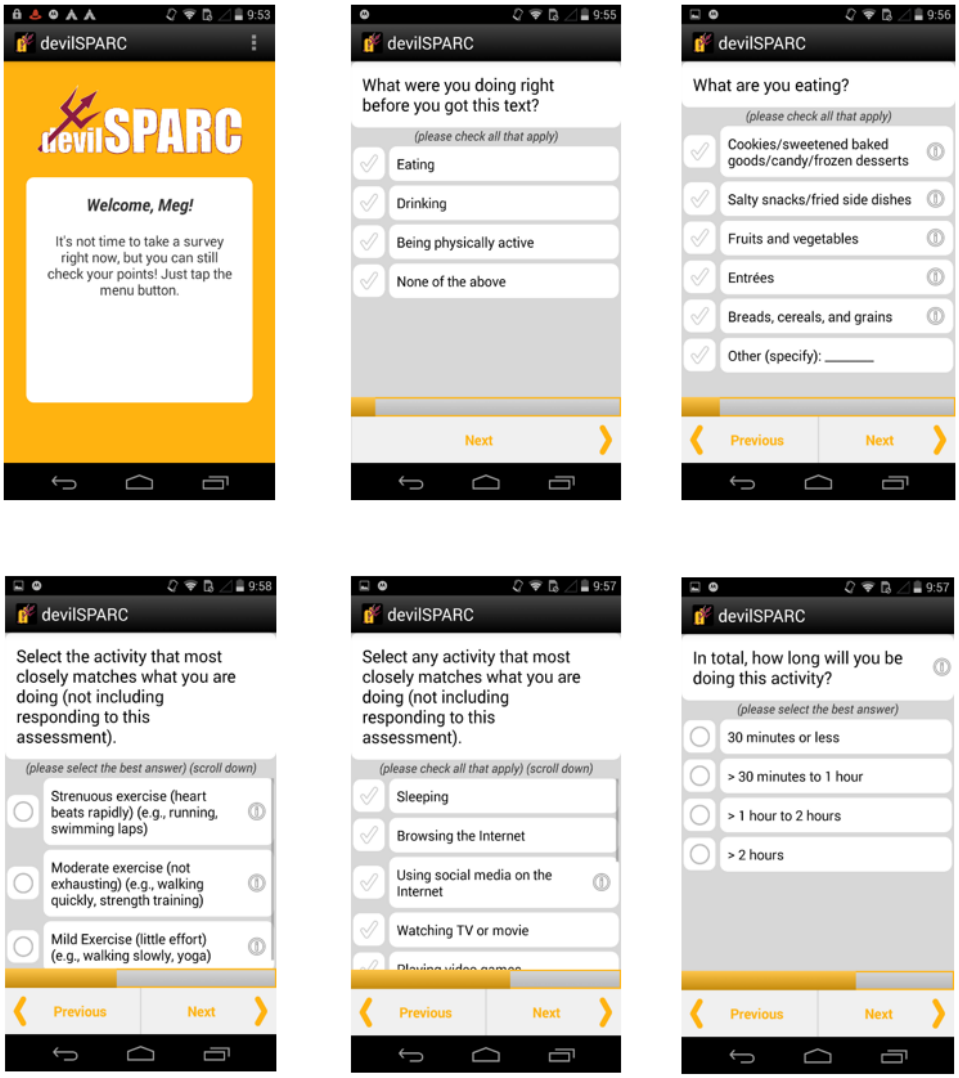
### *Accelerometry*

Actigraph, Inc (model GT3X+) accelerometer devices provided an objective measure for participants' PA. A 60-second epoch for summing counts and the Freedson et al [28] cut points were used to classify PA levels (sedentary <100 counts per minute [CPM], light 100-1951 CPM, moderate 1952-5724 CPM, and vigorous ≥5725 CPM) for 30 minutes prior to and after the mEMA prompt.



XSL•FO
RenderX

**Figure 1.** Screenshots of the eating and physical activity behaviors assessed in the devilSPARC mEMA app.




## Data Analysis

Analyses were specific to dietary and PA data. Given the varied distribution for each of the food choices across both the mEMA and ASA24, comparisons between the mEMA and ASA24 responses at the moment of the mEMA response were not possible. However, because accelerometers capture PA every 60 seconds, direct comparisons between the PA level reported in the mEMA and the PA recorded by the accelerometer at the moment of the mEMA response were possible. The specific analyses are described subsequently.

### *Validating the Mobile Ecological Momentary Assessment Dietary Data*

Participants' data were excluded (n=15) from the analysis for days without (1) mEMA data with at least one food entry and (2) a biologically plausible ASA24 dietary recall. This resulted in an analytical sample of 92 from the 107 participants with ASA information (86.0%). The percentage match between the ASA24 and mEMA at both the daily level, and for time windows around the mEMA (ranging from 6 minutes to 8 hours, in 6-minute increments, on either side of the beginning of each mEMA report) were determined for each food type (sweets; salty snacks/fried side dishes; fruits and vegetables; entrées; and breads, cereals, and grains).

In day-level analyses, for each food type, the denominator for the match rate included the number of times the food type was reported via mEMA during the day; the numerator was the smaller of (1) the number of times the food type was reported in that day's mEMA reports and (2) the corresponding ASA24 count for each participant. For example, if fruits and vegetables were reported four times via mEMA on a given day and three times in that day's ASA24, this was recorded as three matches from a potential four (75% match rate). Conversely, if on a given day fruits and vegetables were reported three times via mEMA and four times in the corresponding ASA24, this would be recorded as three matches out of a potential three (100% match rate). For the time window analysis, the denominator for the match rate represented every mEMA response with the food type recorded; the numerator was the number of times the food type recorded in the mEMA was also in the ASA24 within the given time window.

Chi-square tests were used to determine if match rates differed between males and females, white and nonwhite participants, and participants with and without a Pell grant. Sensitivity and



XSL•FO
RenderX

specificity values were computed for the time windows to determine the impact of increased windows size on the match rates. For the sake of conciseness, positive and negative likelihood ratios are not reported for each food type. Mixed effects logistic regression models with random participant-level intercepts were used to determine how well the endorsement of a food type in mEMA reports could be predicted from the endorsement of the same food type in the ASA24.

### *Validating Mobile Ecological Momentary Assessment Physical Activity Data*

Participants' mEMA responses were excluded (n=8) from the analysis if the accelerometer activity CPM values were zero for the 30 minutes before or after the mEMA response, or if the accelerometer had not been worn for at least 5 hours for the day of the mEMA. This resulted in an analytic sample of 41 from the 49 participants with accelerometer information (84%). Because accelerometer readings showed high minute-to-minute variability, the average accelerometer activity (CPM) value for the 5 minutes prior to the EMA response was used as the measure of accelerometer-derived activity. There were six parameters used to characterize agreement between the PA level as determined by accelerometer activity counts (sedentary, light, moderate, or vigorous, as described in the Measures section) and the PA level reported in the mEMA (sedentary, light, moderate, or strenuous). These parameters were (1) odds ratio: the odds that a participant's accelerometer-derived activity level and mEMA-reported PA level were both at a specific PA level, compared to the odds a participant's accelerometer-derived activity level was at the specified level, but the mEMA-reported PA level was not; (2) match rates: the percentage of times the accelerometer-derived activity level and reported mEMA level were the same for each mEMA level; (3) sensitivity (true positive rate): the percentage of times both the mEMA-reported PA levels and the accelerometer-derived activity level were at the same PA level for each accelerometer-derived PA level; (4) specificity (true negative rate): the percentage of times the mEMA-reported PA levels and the accelerometer-derived activity level were not at the specific PA level for each accelerometer-derived PA level; (5) positive likelihood ratio: the increase in the likelihood that a particular accelerometer-measured activity level (eg, moderate) was achieved, given that the same PA level was reported via mEMA; and (6) negative likelihood ratio: the decrease in the likelihood that a particular accelerometer-measured PA level (eg, moderate) was achieved, given that a *different* PA level was reported via mEMA. To determine if the distribution of the accelerometer-derived activity levels differed systematically with respect to mEMA-reported PA level, two-sample Kolmogorov-Smirnov tests were run. This nonparametric test compares the maximum vertical distance between the cumulative distribution functions of two distributions, with the *P* value corresponding to the probability that the distributions are the same (ie, small *P* values indicate greater discrepancy between the forms of the distributions). Mixed effects linear regression models with random participant-level intercepts were used to determine how well mEMA-reported PA categories predicted log transformed accelerometer-derived activity levels. All analyses were conducted using R statistical software version 3.2.3 (R Foundation for Statistical Computing, Vienna, Austria). Statistical significance was determined at $P<.05$.

## Results

Data from 92 participants were used in analyses examining validity of mEMA-reported dietary behavior (age: mean 18.83, SD 0.61 years; female: 67/92, 67%), and data from 41 participants were used in analyses aimed at examining the validity of mEMA-reported PA (age: mean 18.72, SD 0.50 years; female: 30/41, 73%; see Table 1).

### Dietary Validation

A total of 272 mEMA prompts and 607 ASA24 eating instances from 163 participant days were analyzed, including those from 17 participants who provided three days of ASA24 recall data, 37 participants who provided two days of data, and 38 participants who provided one day of data. Entrée was the most common food type reported in the mEMA (121/272, 44.5%), but was the least-reported food type reported in the ASA24 (294/607, 48.4%; see Table 2). At the day level, the percentage of occasions when a food type reported in the mEMA was also reported in the ASA24 ranged from 79% (95/121 entrées reported in the mEMA matched to ASA) to 94% (64/68 fruit and vegetables reported in the mEMA matched to ASA).



XSL•FO
RenderX

**Table 1.** Participant demographics in mEMA diet validation and PA validation.

| Demographic variable | Diet validation (n=92) | PA validation (n=41) |
|---|---|---|
| **Gender, n (%)** | | |
| Male | 30 (33) | 11 (27) |
| Female | 62 (67) | 30 (73) |
| Age (years), mean (SD) | 18.83 (0.61) | 18.72 (0.50) |
| **Race/ethnicity, n (%)** | | |
| White only | 54 (59) | 21 (51) |
| Black only | 4 (4) | 4 (10) |
| Mixed/other | 17 (18) | 5 (12) |
| Hispanic | 17 (18) | 11 (27) |
| Pell grant status (yes), n (%) | 29 (32) | 18 (44) |
| **Major, n (%)** | | |
| Humanities | 14 (15) | 8 (20) |
| Natural sciences | 44 (48) | 18 (44) |
| Social sciences | 22 (24) | 9 (22) |
| Other | 12 (13) | 6 (15) |
| **Year in college, n (%)** | | |
| First | 86 (93) | 41 (100) |
| Second | 2 (2) | 0 (0) |
| Third | 4 (4) | 0 (0) |

**Table 2.** Number and percentage of times each food type was observed at the daily level for the mEMA and ASA24, and match rate at the daily level for each food type.

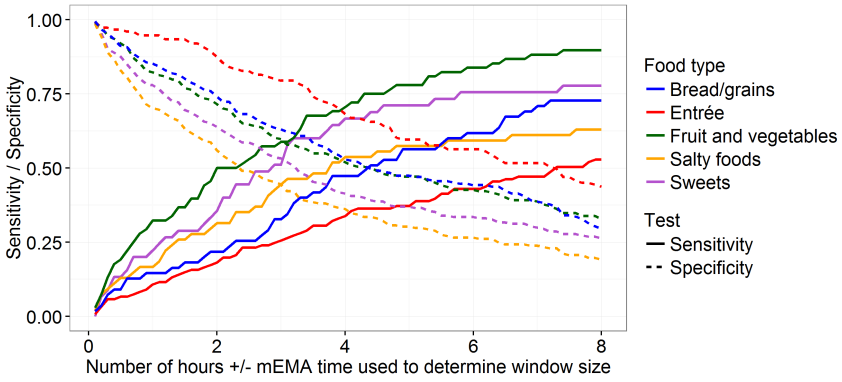
| Self-reported food group | mEMA, n (%) (n=272) | ASA24, n (%) (n=607) | Match rate (%) |
|---|---|---|---|
| Bread/grains | 55 (20) | 392 (65) | 89 |
| Entrée | 121 (44) | 294 (48) | 79 |
| Fruit and vegetables | 68 (25) | 347 (57) | 94 |
| Salty foods | 54 (20) | 426 (70) | 80 |
| Sweets | 45 (17) | 404 (67) | 91 |

Increased times on either side of the mEMA response resulted in increased specificity and decreased sensitivity (see Figure 2). Although participants reported the same foods in the mEMA and the ASA24, they were not accurate with the time they reported foods in the ASA24. In general, entrées had the lowest match rate across all the time windows, and fruits and vegetables had the highest match rate. There was no significant difference between the mEMA and ASA24 match rates by gender, race/ethnicity, or Pell grant status (data not shown). No significant associations between mEMA-reported food types and ASA24 reports were observed in the mixed model results (data not shown).



XSL•FO
RenderX

**Figure 2.** Sensitivity and specificity for each food type with increasing time window size.




## Physical Activity Validation

A total of 694 mEMA surveys with valid accelerometer values across the 41 participants were included in the analysis. Table 3 presents the frequency and match rate of the mEMA activity and corresponding accelerometer-derived activity. Sedentary or light PA were the most often mEMA-reported activity levels (628/694, 90.5% of mEMA reports). Approximately 95% (656/694) of the accelerometer-derived activity levels corresponded with sedentary or light PA. Of the 26 mEMA responses reporting vigorous PA, only one participant's accelerometer-derived activity levels indicated vigorous intensity.

**Table 3.** Cross-tabulation of frequencies of mEMA-reported and accelerometer-derived physical activity levels in 41 participants (n=694 mEMA reports).

| mEMA-reported activity level | Accelerometer-derived activity level, n | | | | Total of mEMA reports, n (%) |
|---|---|---|---|---|---|
| | Sedentary | Light | Moderate | Vigorous | |
| Sedentary | 340 | 209 | 16 | 0 | 565 (81.4) |
| Light | 19 | 37 | 7 | 0 | 63 (9.1) |
| Moderate | 11 | 20 | 9 | 0 | 40 (5.8) |
| Vigorous | 2 | 18 | 5 | 1 | 26 (3.8) |
| Total of accelerometer counts, n (%) | 372 (53.6) | 284 (40.9) | 37 (5.3) | 1 (0.1) | 694 (100) |

The odds of a participant having their accelerometer-derived activity level match their reported PA level were significant for mEMA-reported sedentary PA (OR 4.69, 95% CI 3.00-7.32), light PA (OR 2.27, 95% CI 1.32-3.88), and moderate PA (OR 6.30, 95% CI 2.65-14.95) Due to only one participant having vigorous accelerometer values, odds were not computed for vigorous activity. The match rates were highest for mEMA-reported sedentary and light PA (340/565, 60.3% and 37/63, 58.7%, respectively), and lowest for moderate PA (9/40, 22.5%) and vigorous PA (1/26, 3.8%).

We also conducted sensitivity and specificity between mEMA-reported activities and accelerometer-derived activities, for each respective PA intensity level. Specificity and positive likelihood ratio values were lower for mEMA-reported sedentary PA (specificity=30%, positive likelihood ratio=1.31) than mEMA-reported light (specificity=94%, positive likelihood ratio=2.05), moderate (specificity=95%, positive likelihood ratio=5.16), and vigorous (specificity=96%, positive likelihood ratio=27.72) PA. Sensitivity values were highest for mEMA-reported sedentary (91%) and vigorous PA (100%), and lowest for mEMA-reported light (13%) and moderate PA (24%); negative likelihood ratio values were lowest for mEMA-reported sedentary (0.29) and vigorous PA (0.00), and highest for mEMA-reported light (0.93) and moderate PA (0.79).

The difference in participants' average accelerometer-derived activity levels was not consistent across the mEMA levels. For example, as illustrated in Figure 3, for eight of 21 participants who reported sedentary and light PA via mEMA, the average accelerometer-derived activity level was *lower* on occasions when they reported light PA than on occasions when they reported being sedentary. Similarly, for six of the 14 participants who reported both light and moderate PA, on occasions when moderate PA was reported average, accelerometer-derived activity levels were lower than occasions when light PA was reported.

To examine the differences in distributions of accelerometer-derived activity levels for different reported mEMA PA levels, we conducted Kolmogorov-Smirnov tests. Results indicated a significant difference in the distribution of the accelerometer-derived activity counts for mEMA-reported sedentary and nonsedentary PA levels ($P$<.001; Table 3). The difference between the distribution of the accelerometer-derived activity counts for mEMA-reported light and vigorous PA was also significant ($P$=.02). No difference between the distribution of the accelerometer-derived activity counts were seen between mEMA-reported light and moderate PA occasions, or moderate and strenuous PA occasions.



XSL•FO
RenderX

**Table 4.** Kolmogorov-Smirnov results examining whether the accelerometer-derived activity count distributions for each pair of mEMA levels could be from the same distribution.

| Activity levels as reported in mEMA | Activity levels as reported in mEMA, *P* | | |
|---|---|---|---|
| | Light (n=63) | Moderate (n=40) | Strenuous (n=26) |
| Sedentary (n=565) | <.001 | <.001 | <.001 |
| Light (n=63) | — | .13 | .02 |
| Moderate (n=40) | — | — | .26 |

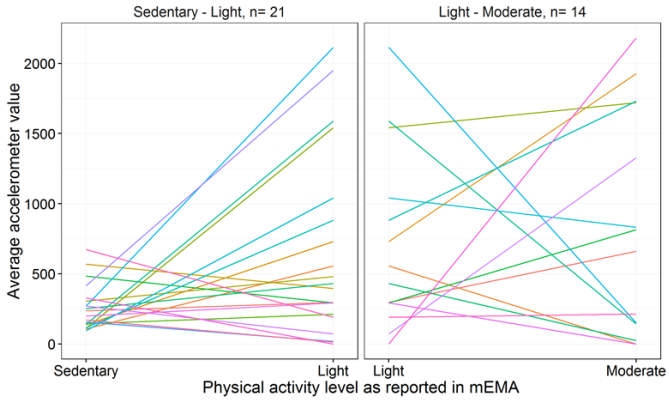
To more closely examine the association between mEMA-reported intensity of PA to accelerometer-measured levels, we estimated differences in distributions of logged accelerometer activity count values for pairs of mEMA-reported PA levels using mixed linear regression models (with repeated observations nested within participants) (Table 5). There was a significant (*P*<001) difference between logged activity count values for mEMA-reported sedentary and nonsedentary accelerometer occasions, but no significant difference in logged counts across mEMA-reported light, moderate, and strenuous PA occasions (*P*=.84, *P*=.05, and *P*=.10, respectively). For example, when comparing mEMA-reported sedentary versus light PA occasions, activity counts were higher for mEMA-reported light PA occasions than for sedentary occasions (*P*<.001). The estimated average increases in logged accelerometer activity counts between sedentary and nonsedentary occasions were 1.71, 1.81, and 2.79 for light, moderate, and vigorous PA, respectively, corresponding to differences of 178, 201, and 603 CPM, respectively, in raw count values.

**Table 5.** Estimated differences, 95% confidence intervals, and *P* values for pairwise comparisons of logged accelerometer activity counts between mEMA-reported PA levels.[a]

| mEMA-reported PA level | mEMA-reported PA level | | | | | |
|---|---|---|---|---|---|---|
| | Light | | Moderate | | Vigorous | |
| | Diff (95% CI) | *P* | Diff (95% CI) | *P* | Diff (95% CI) | *P* |
| Sedentary | 1.71 (1.09-2.33) | <.001 | 1.81 (1.04-2.57) | <.001 | 2.79 (1.85-3.73) | <.001 |
| Light | – | | 0.10 (–0.85-1.04) | .84 | 1.08 (–0.01-2.18) | .05 |
| Moderate | – | | – | | 0.99 (–0.19-2.17) | .10 |

[a] Estimates from mixed models adjusted for nonindependence of repeated within-person observations.

**Figure 3.** Within-person difference in accelerometer values by reported physical activity levels.




## Discussion

### How Well Does the devilSPARC App Measure Eating and Physical Activity Behaviors?

This study assessed the validity of the devilSPARC mEMA app as a tool for assessing eating and PA behaviors among young adults compared to online dietary recall and accelerometry methodologies. Diet and PA assessment methods commonly used in current research settings often require high levels of cost and personal effort for participants. Few objective assessments of dietary quality and intake are available; self-report remains the norm in observational studies. Objective assessments of PA tend to use expensive devices. The mEMA had high match rates with day-level reported dietary intake as measured by 24-hour recall. For PA behaviors, mEMA reports differentiated sedentary from nonsedentary activity, but these reports did not accurately distinguish among objectively measured PA levels. These findings suggest that the devilSPARC mEMA app had relatively high criterion validity with food choices and for distinguishing between sedentary versus nonsedentary activity.



XSL•FO
RenderX

Research has demonstrated that methods for dietary recall are subject to significant compliance, self-reporting, and recall errors [3,28]. These analyses excluded several participants because of biologically implausible values in the 24-hour dietary recall measure. With the exception of one, all exclusions were a result of participants reporting daily intakes of less than 500 kilocalories. Anecdotally, many participants reported frustration with the functionality of the ASA24 website and the amount of time it took to complete the recall. As with findings from other studies, we expect that the 24-hour recall data reported here underrepresents dietary intake and misestimates the time at which participants consumed food [29-32]. The potential lack of adherence to the 24-hour recall protocol may explain why the sensitivity for food choices increased over time, from approximately 70% for 8-hour windows surrounding the time at which a given food was reported on the mEMA to 10% for half-hour windows (ie, lower match in shorter window). Although still relatively high, the measure of entrées showed the lowest match rate at 79% between the mEMA and the ASA24; this is likely because of the lack of specificity of what participants perceived as an entrée. Our results demonstrated that with just a few questions, devilSPARC mEMA may be able to assess food choices with significantly less burden than the self-administered 24-hour dietary recall for each eating occasion, particularly given there was relatively high construct validity.

Reports of light, moderate, and vigorous levels of PA from the mEMA did not correspond to intensity of PA as measured objectively through accelerometry. The proportion of participants reporting an activity level that corresponded to the accelerometer decreased with increasing PA level. Social desirability and/or perception biases may be at play with these results. Social desirability is often related to over-reporting of activity duration and intensity [1,33,34]. Other research has also reported that the percent agreement in validating mEMA is highest for sedentary activity [35]. In addition, individuals who are heavier or are less fit may perceive an activity to be more intense due to increased respiration and heart rates [36]. Because established accelerometer activity cut points do not take into account body weight or current fitness level, energy expenditure may vary across individuals who are engaging in the same amount of objectively measured PA.

### Limitations

To our knowledge, this is the first study to validate a mEMA tool assessing eating and PA behaviors among young people; however, several limitations should be considered. The devilSPARC app did not assess quantity of foods or specific details of the foods (ie, brand or type), as this would have added to the response burden; the devilSPARC tool assessed broad behaviors and was therefore not able to yield information about total caloric intake, macronutrients, or micronutrients. Given the relatively equal distribution of food choices captured, we were able to assess a variety of commonly consumed foods, including healthy and unhealthy food choices, for young adults in college. The devilSPARC mEMA tool was designed to assess behaviors in the moment; as such, it could not represent total dietary intake or total PA. In addition, because the mEMA and the 24-hour recall relied on self-reports, participants' reporting biases and idiosyncratic interpretations of mEMA questions could have increased measurement error. Despite verbal and written directions to wear the accelerometer at all times except when swimming and bathing, there is a possibility that participants removed the accelerometer when participating in vigorous activities, such as contact sports, which would have improved our match rate between the mEMA and more vigorous activities. Also, although the sample was relatively diverse in terms of race/ethnicity, almost all participants were college freshmen; thus, these findings may not be generalizable beyond young adult populations.

### Conclusions

This new mEMA tool is valid for assessing eating behaviors and the presence of PA. With very brief surveys spaced through the day, this mEMA tool may reduce participant burden as compared to 24-hour dietary recall or PA recall instruments. The mEMA builds on previous measures of assessing eating and PA, including a wide range of foods.

## Acknowledgments

A special acknowledgment goes to our masterful programmers, Kevin Hollingshead and John Yu, who helped to create and test the devilSPARC app. We would like to thank the students for taking the time to participate in this study and the devilSPARC research team for their assistance in collecting the data, including Kara Skelton. This study was supported by the NIH Common Fund from the Office of the Director and the Office of Behavioral and Social Sciences Research, grant number 1DP5OD017910-01 (PI: M Bruening). The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

## Conflicts of Interest

None declared.

## References

1. Brenner PS, DeLamater JD. Social desirability bias in self-reports of physical activity: is an exercise identity the culprit? Soc Indic Res 2014;117(2):489-504. [doi: 10.1007/s11205-013-0359-y]
2. Spook JE, Paulussen T, Kok G, Van Empelen P. Monitoring dietary intake and physical activity electronically: feasibility, usability, and ecological validity of a mobile-based ecological momentary assessment tool. J Med Internet Res 2013 Sep 24;15(9):e214 [FREE Full text] [doi: 10.2196/jmir.2617] [Medline: 24067298]



XSL•FO
RenderX

3. Livingstone MB, Robson PJ, Wallace JM. Issues in dietary intake assessment of children and adolescents. Br J Nutr 2004 Oct;92 Suppl 2:S213-S222. [Medline: 15522159]
4. Sallis JF, Saelens BE. Assessment of physical activity by self-report: status, limitations, and future directions. Res Q Exerc Sport 2000 Jun;71 Suppl 2:1-14. [doi: 10.1080/02701367.2000.11082780] [Medline: 25680007]
5. Chinapaw MJ, Mokkink LB, van Poppel MN, van Mechelen W, Terwee CB. Physical activity questionnaires for youth: a systematic review of measurement properties. Sports Med 2010 Jul 1;40(7):539-563. [doi: 10.2165/11530770-000000000-00000] [Medline: 20545380]
6. Shephard RJ. Limits to the measurement of habitual physical activity by questionnaires. Br J Sports Med 2003 Jun;37(3):197-206; discussion 206 [FREE Full text] [Medline: 12782543]
7. Shiffman S, Stone AA, Hufford MR. Ecological momentary assessment. Annu Rev Clin Psychol 2008 Apr;4:1-32. [doi: 10.1146/annurev.clinpsy.3.022806.091415] [Medline: 18509902]
8. Stone AA, Shiffman S, Atienza AA, Nebeling L. The Science of Real-Time Data Capture. New York: Oxford University Press, Inc; 2007.
9. Daugherty BL, Schap TE, Ettienne-Gittens R, Zhu FM, Bosch M, Delp EJ, et al. Novel technologies for assessing dietary intake: evaluating the usability of a mobile telephone food record among adults and adolescents. J Med Internet Res 2012 Apr 13;14(2):e58 [FREE Full text] [doi: 10.2196/jmir.1967] [Medline: 22504018]
10. Kuntsche E, Labhart F. Using personal cell phones for ecological momentary assessment. Eur Psychol 2013 Apr 08;18(1):3-11. [doi: 10.1027/1016-9040/a000127]
11. Schüz N, Cianchi J, Shiffman S, Ferguson SG. Novel technologies to study smoking behavior: current developments in ecological momentary assessment. Curr Addict Rep 2015 Mar;2(1):8-14. [doi: 10.1007/s40429-015-0039-x]
12. Dunton GF, Intille SS, Wolch J, Pentz MA. Children's perceptions of physical activity environments captured through ecological momentary assessment: a validation study. Prev Med 2012 Aug;55(2):119-121. [doi: 10.1016/j.ypmed.2012.05.015] [Medline: 22659225]
13. Dunton GF, Dzubur E, Kawabata K, Yanez B, Bo B, Intille S. Development of a smartphone application to measure physical activity using sensor-assisted self-report. Front Public Health 2014 Feb 28;2:12 [FREE Full text] [doi: 10.3389/fpubh.2014.00012] [Medline: 24616888]
14. Phillips MM, Phillips KT, Lalonde TL, Dykema KR. Feasibility of text messaging for ecological momentary assessment of marijuana use in college students. Psychol Assess 2014 Sep;26(3):947-957. [doi: 10.1037/a0036612] [Medline: 24749751]
15. Liao Y, Skelton K, Dunton G, Bruening M. A systematic review of methods and procedures used in ecological momentary assessments of diet and physical activity research in youth: an adapted STROBE Checklist for Reporting EMA Studies (CREMAS). J Med Internet Res 2016 Jun 21;18(6):e151 [FREE Full text] [doi: 10.2196/jmir.4954] [Medline: 27328833]
16. Nelson MC, Story M, Larson NI, Neumark-Sztainer D, Lytle LA. Emerging adulthood and college-aged youth: an overlooked age for weight-related behavior change. Obesity (Silver Spring) 2008 Oct;16(10):2205-2211. [doi: 10.1038/oby.2008.365] [Medline: 18719665]
17. Loria CM, Signore C, Arteaga SS. The need for targeted weight-control approaches in young women and men. Am J Prev Med 2010 Feb;38(2):233-235 [FREE Full text] [doi: 10.1016/j.amepre.2009.11.001] [Medline: 20117581]
18. Lytle LA, Svetkey LP, Patrick K, Belle SH, Fernandez ID, Jakicic JM, et al. The EARLY trials: a consortium of studies targeting weight control in young adults. Transl Behav Med 2014 Sep;4(3):304-313 [FREE Full text] [doi: 10.1007/s13142-014-0252-5] [Medline: 25264469]
19. Palfrey J, Gasser U. Born Digital: Understanding the First Generation of Digital Natives. New York: Basic Books; 2008.
20. Berry DM, Bass CP. Successfully recruiting, surveying, and retaining college students: a description of methods for the Risk, Religiosity, and Emerging Adulthood Study. Res Nurs Health 2012 Dec;35(6):659-670. [doi: 10.1002/nur.21498] [Medline: 22753170]
21. Laska MN, Graham D, Moe SG, Lytle L, Fulkerson J. Situational characteristics of young adults' eating occasions: a real-time data collection using Personal Digital Assistants. Public Health Nutr 2011 Mar;14(3):472-479. [doi: 10.1017/S1368980010003186] [Medline: 21138611]
22. Godin G, Shephard RJ. A simple method to assess exercise behavior in the community. Can J Appl Sport Sci 1985 Sep;10(3):141-146. [Medline: 4053261]
23. Moshfegh AJ, Rhodes DG, Baer DJ, Murayi T, Clemens JC, Rumpler WV, et al. The US Department of Agriculture Automated Multiple-Pass Method reduces bias in the collection of energy intakes. Am J Clin Nutr 2008 Aug;88(2):324-332 [FREE Full text] [Medline: 18689367]
24. Phillips SM, Bandini LG, Naumova EN, Cyr H, Colclough S, Dietz WH, et al. Energy-dense snack food intake in adolescence: longitudinal relationship to weight and fatness. Obes Res 2004 Mar;12(3):461-472 [FREE Full text] [doi: 10.1038/oby.2004.52] [Medline: 15044663]
25. Laska MN, Hearst MO, Forsyth A, Pasch KE, Lytle L. Neighbourhood food environments: are they associated with adolescent dietary intake, food purchases and weight status? Public Health Nutr 2010 Nov;13(11):1757-1763 [FREE Full text] [doi: 10.1017/S1368980010001564] [Medline: 20529405]




XSL•FO
RenderX

26. Berkey CS, Rockett HR, Field AE, Gillman MW, Frazier AL, Camargo CA, et al. Activity, dietary intake, and weight changes in a longitudinal study of preadolescent and adolescent boys and girls. Pediatrics 2000 Apr;105(4):E56. [Medline: 10742377]
27. Eaton DK, Olsen EO, Brener ND, Scanlon KS, Kim SA, Demissie Z, et al. A comparison of fruit and vegetable intake estimates from three survey question sets to estimates from 24-hour dietary recall interviews. J Acad Nutr Diet 2013 Sep;113(9):1165-1174 [FREE Full text] [doi: 10.1016/j.jand.2013.05.013] [Medline: 23871104]
28. Freedson PS, Melanson E, Sirard J. Calibration of the Computer Science and Applications, Inc accelerometer. Med Sci Sports Exerc 1998 May;30(5):777-781. [Medline: 9588623]
29. Livingstone MB, Black AE. Markers of the validity of reported energy intake. J Nutr 2003 Mar;133 Suppl 3:895S-920S [FREE Full text] [Medline: 12612176]
30. Vereecken CA, Covents M, Sichert-Hellert W, Alvira JM, Le Donna C, De Henauw S, et al. Development and evaluation of a self-administered computerized 24-h dietary recall method for adolescents in Europe. Int J Obes (Lond) 2008 Nov;32 Suppl 5:S26-S34. [doi: 10.1038/ijo.2008.180] [Medline: 19011650]
31. Ma Y, Olendzki BC, Pagoto SL, Hurley TG, Magner RP, Ockene IS, et al. Number of 24-hour diet recalls needed to estimate energy intake. Ann Epidemiol 2009 Aug;19(8):553-559 [FREE Full text] [doi: 10.1016/j.annepidem.2009.04.010] [Medline: 19576535]
32. De Keyzer W, Huybrechts I, De Vriendt V, Vandevijvere S, Slimani N, Van Oyen H, et al. Repeated 24-hour recalls versus dietary records for estimating nutrient intakes in a national food consumption survey. Food Nutr Res 2011;55:7307 [FREE Full text] [doi: 10.3402/fnr.v55i0.7307] [Medline: 22084625]
33. Adams SA, Matthews CE, Ebbeling CB, Moore CG, Cunningham JE, Fulton J, et al. The effect of social desirability and social approval on self-reports of physical activity. Am J Epidemiol 2005 Feb 15;161(4):389-398 [FREE Full text] [doi: 10.1093/aje/kwi054] [Medline: 15692083]
34. Motl RW, McAuley E, DiStefano C. Is social desirability associated with self-reported physical activity? Prev Med 2005 Jun;40(6):735-739. [doi: 10.1016/j.ypmed.2004.09.016] [Medline: 15850873]
35. Dunton GF, Whalen CK, Jamner LD, Henker B, Floro JN. Using ecologic momentary assessment to measure physical activity during adolescence. Am J Prev Med 2005 Nov;29(4):281-287. [doi: 10.1016/j.amepre.2005.07.020] [Medline: 16242590]
36. Ekkekakis P, Lind E. Exercise does not feel the same when you are overweight: the impact of self-selected and imposed intensity on affect and exertion. Int J Obes (Lond) 2006 Apr;30(4):652-660. [doi: 10.1038/sj.ijo.0803052] [Medline: 16130028]


## Abbreviations

**API:** application programming interfaces
**CPM:** counts per minute
**EMA:** ecological momentary assessment
**mEMA:** mobile-based ecological momentary assessment
**NIH:** National Institutes of Health
**PA:** physical activity
**SMS:** short message service
**SPARC:** Social impact of Physical Activity and nutRition in College
**USDA:** US Department of Agriculture


*Edited by G Eysenbach; submitted 11.05.16; peer-reviewed by G Kok, K Phillips; comments to author 04.06.16; revised version received 19.06.16; accepted 21.06.16; published 27.07.16.*

*Please cite as:*
*Bruening M, van Woerden I, Todd M, Brennhofer S, Laska MN, Dunton G*
*A Mobile Ecological Momentary Assessment Tool (devilSPARC) for Nutrition and Physical Activity Behaviors in College Students: A Validation Study*
*J Med Internet Res 2016;18(7):e209*
*URL: http://www.jmir.org/2016/7/e209/*
*doi:10.2196/jmir.5969*
*PMID:27465701*


©Meg Bruening, Irene van Woerden, Michael Todd, Stephanie Brennhofer, Melissa N. Laska, Genevieve Dunton. Originally published in the Journal of Medical Internet Research (http://www.jmir.org), 27.07.2016. This is an open-access article distributed



XSL•FO
RenderX

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



XSL•FO
RenderX

Original Paper

# Follow #eHealth2011: Measuring the Role and Effectiveness of Online and Social Media in Increasing the Outreach of a Scientific Conference

Marcel Winandy[1], PhD; Patty Kostkova[2*], PhD; Ed de Quincey[3*], PhD; Connie St Louis[4], PhD; Martin Szomszor[5], PhD

[1]Horst Goertz Institute for IT-Security, Bochum, Germany
[2]University College London, London, United Kingdom
[3]Keele University, School of Computing and Mathematics, Keele, United Kingdom
[4]City University, London, United Kingdom
[5]Data Science, London, United Kingdom
[*]these authors contributed equally

**Corresponding Author:**
Patty Kostkova, PhD
University College London
Gower St
London,
United Kingdom
Phone: 44 7813787366
Fax: 44 78137873666
Email: p.kostkova@ucl.ac.uk

## Abstract

**Background:** Social media promotion is increasingly adopted by organizers of industry and academic events; however, the success of social media strategies is rarely questioned or the real impact scientifically analyzed.

**Objective:** We propose a framework that defines and analyses the impact, outreach, and effectiveness of social media for event promotion and research dissemination to participants of a scientific event as well as to the virtual audience through the Web.

**Methods:** Online communication channels Twitter, Facebook, Flickr, and a Liveblog were trialed and their impact measured on outreach during five phases of an eHealth conference: the setup, active and last-minute promotion phases before the conference, the actual event, and after the conference.

**Results:** Planned outreach through online channels and social media before and during the event reached an audience several magnitudes larger in size than would have been possible using traditional means. In the particular case of eHealth 2011, the outreach using traditional means would have been 74 attendees plus 23 extra as sold proceedings and the number of downloaded articles from the online proceedings (4107 until October 2013). The audience for the conference reached via online channels and social media was estimated at more than 5300 in total during the event. The role of Twitter for promotion before the event was complemented by an increased usage of the website and Facebook during the event followed by a sharp increase of views of posters on Flickr after the event.

**Conclusions:** Although our case study is focused on a particular audience around eHealth 2011, our framework provides a template for redefining "audience" and outreach of events, merging traditional physical and virtual communities and providing an outline on how these could be successfully reached in clearly defined event phases.

***(J Med Internet Res 2016;18(7):e191)*** doi:10.2196/jmir.4480

**KEYWORDS**

social media; social media networks; Web conferencing; marketing



XSL•FO
RenderX

## Introduction

Measuring the impact or influence of a particular scientific or business event is an important part of evaluating its success and the effectiveness of its promotion. Although social media promotion is a "must" for most commercial and academic events, little interest has been given to defining new audiences participating virtually and physically as well as analyzing the impact and outreach of all individual social media channels used in promotion and scientific outreach. Traditionally, the impact of a scientific conference has been measured by the number of attendees and the number and quality of publications (in terms of acceptance rate and citations). These measures are based on the traditional means of communication with physical communities: face-to-face meetings and printed media. However, in an increasingly widely connected world, the use of social media and novel online channels spanning the traditional physical and virtual divide has revolutionized communication outreach, community engagement, and the overall impact of a scientific conference that can embrace and utilize the new media channels.

In this context, it is important to know the role and effectiveness of online and social media channels in engaging a community. There is a vast amount of research around the usage and effects of social media (eg, [1-4]). Among the factors that most research focuses on are analyzing the dynamics of social networks, information diffusion and propagation, users influence, and attention. However, little attention has been given to investigating the impact of a social network on a physical community, around a single topic, over an extended period of time, and how intensive, face-to-face interaction and virtual socio-patterns at a conference affect the size and constituency of the virtual network. Secondly, most research has looked into an isolated social network or media, such as Twitter or Facebook, but investigating the role of these channels in creating and engaging a community has not been addressed. Finally, most existing research investigates a snapshot image of the entire social network (one-way mining data from the social network, eg, Twitter). In contrast, we have conducted a longitudinal study over 6 months (two-way sending data to and mining from social networks).

In this paper, we make the following contributions:

We define a framework of (1) media channels and their impact factors as well as (2) establishment of longitudinal phases and analysis measures. With the first part of the framework, we aim to study the relationship between an online and real community, and we compare traditional and new impact factors of the outreach of a real-world scientific event. The second part of the framework aims to analyze how successful each media channel is throughout different phases of planning and running an event.

We evaluate our framework in a case study of a real event (ie, the eHealth 2011 conference), which took place in 2011 in Malaga, Spain. We present a detailed analysis of the data we collected through a longitudinal evaluation over different phases to determine the outreach of each media channel and how to calculate the detailed activity on creating and engaging an online community around a conference.

Finally, we discuss the results of our case study and aim to answer the question of which role is best suited for each media channel before, during, and after a conference.

The objective of our research is to investigate methods to determine the impact of different media channels on a real event over traditional research event dissemination methods. To this end, we define and suggest a strategy for promotional phases before, during, and after the event, and an outreach score as a measure to determine the impact.

### Related Work

The availability of Twitter datasets has created a rapid increase in research projects across a range of domains investigating influence, propagation, information diffusion, and social network topology. There has been some interest in investigating the role of social media to improve user experience and engagement with conferences, for example, mobile phone apps such as Conference Navigator presented at UMAP 2011 [5] and conference organizing apps using social media [6]. However, neither work addresses the use of multiple channels of outreach beyond the event itself.

Research on the dynamics and influence of the network itself primarily address various issues of creating influence and activity versus passivity of users to post, retweet, and mention. Influence and passivity scores investigated by Romero et al [2], Meeder et al [7], and Bruns et al [8] introduced methods to retrospectively determine follower growth on Twitter accounts. In contrast, we collect the data on the number of followers of our account on a regular basis and we also observe the numbers of users who unfollow. Unlike most studies that investigate a snapshot of the social network, our research looks into a longitudinal community behavior and long-term impact. Golbeck [9] conducted a longitudinal study of membership growth in various social networks and observed a linear increase in most cases. However, her research looks into general growth of an entire social media network. Instead, our research targets a specific community across various media around a single event. Russo et al [3] presented a longitudinal study investigating a relationship between tagging and attention on a variety of social networks. Cosley and Lan [4] studied social influence using Wikipedia. Although these works demonstrate that people with a high density of interconnection actually share less information about the content and context.

An attempt to investigate the correlation between social networks and real networks was investigated by Tugkeci [10], who looked at 617 users using qualitative and quantitative methods. This study revealed that there was no difference in the number of offline friends between those who made new friends online and those who did not. However, the aim of our study is to investigate a real community of professionals with interest in eHealth who had an opportunity to meet face-to-face at the conference. Thus, the online network converged to a face-to-face interaction.

Research into dynamics and activity of user influence on Twitter has also been flourishing. For example, Cha et al [1] analyzed across three measures: indegree (ie, number of followers), retweets, and mentions. They analyzed a large dataset over



XSL•FO
RenderX

(almost all) Twitter users to investigate the influence of single user types and how this influence can remain constant across different topics and over time. Although we also use indegree, retweets, and mentions among our measures for outreach, we do not aim to investigate a single influence but a community as a whole, and we aim for high-density spider networks around a single topic. Although Cha et al and others analyze only a snapshot of activity of all users at the time of their data crawling, we analyze the temporal change of the users' network around an event.

Finally, there is also research focusing on social media usage in academic conferences. For example, the study by McKendrick et al [11] demonstrated the use of social media at health care conferences. They analyzed and categorized tweets that were posted before, during, and after the event. Wen et al [12] analyzed the usage of Twitter for several academic conferences over a time period of 5 years. Although both works show new insights into the usage and network structures of Twitter around conferences and their change over time. Moreover, Wen et al focus on datasets 2 weeks around each event, whereas we not only define larger longitudinal phases that range from several months before the conference up to weeks afterwards, but we also generate data by ourselves, turning the conference audience into a real-world laboratory. Other related areas of research include analysis of why and how people in particular use Twitter during academic conferences [13-18]. A different research direction—measuring interaction socio-patterns and close proximity interaction at a scientific conference using radio-frequency identification (RFID) sensors—was conducted by Barrat et al [19] and Szomszor et al [20], but these studies focused on the physical interaction during the event rather than the wider outreach of scientific outputs. In contrast, we aim to look at what social media channels are best suited to increase outreach and when. Being more closely aligned to our goals, a few case studies [21] aim to examine how to use various social media to increase outreach of scientific conferences. But, most of these are oversimplified and concentrate on one particular channel. Instead, we provide a framework of how to integrate several channels with their specific roles in longitudinal phases and to measure their outreach in terms of certain impact factors. This is achieved by addressing the following questions:

1. What is the relationship between an online and real community in terms of coverage and overlap and what are the impact factors of different (social) media channels through which the communities are built?

2. How successful is each social media channel in the phases of planning and running a real event (ie, a conference)?

3. What is the overall outreach and how to calculate the detailed activity in creating and engaging an online community around a conference?

# Methods

## Framework

In our framework, we considered the following online media channels and their role in establishing extended outreach and community growth:

1. Twitter to provide general, dynamic information about the conference, promote the event, link to relevant news and other information, define a dedicated hashtag for the conference, and actively establish a community of potentially interested followers;

2. Facebook to provide general, dynamic information about the conference, promote the event, link to relevant news and other information, and connect to the Twitter account;

3. Flickr to create a dedicated group for the conference and post images and abstracts of all posters of the conference;

4. Liveblog to provide live blog messages during the event and link in all tweets from Twitter that used the conference hashtag;

5. Website to provide general (static) information about the conference and provide links to the submission system and to all other media channels; and

6. email to send call-for-papers and call-for-participations to mailing lists.

Each of these channels has a different impact on the outreach of the conference. Among the "traditional" impact measures are the number of conference attendees, the number of printed proceedings that are sold, and the number of papers downloaded as electronic versions. These traditional impact measures are often related to the physical community of the conference and the associated traditional media channels. In contrast, social media channels offer a new set of impact measures. These include the number of followers of a social media account, the number of tweets or posts that are related to the conference, and (compared to downloads of papers) the number of page visits on the Flickr group that hosts the images and abstracts of the posters presented at the conference. Multimedia Appendix 1 summarizes the impact measures that we determined based on the corresponding media channels.

### *Establishment and Measurement*

To measure the outreach of the event with the new impact measures, we first set up a number of media channels for the event. Next, we defined five phases along the timeline of the event, ranging from early time before the event up to a time after the event. As such, we could analyze the growth and change of the virtual and physical communities around this real event in a longitudinal manner.

### *Longitudinal Phases*

The five longitudinal phases were oriented around the Fourth Institute for Computer Sciences, Social Informatics and Telecommunications Engineering (ICST) International Conference on eHealth (eHealth 2011), which took place in Malaga, Spain, on November 21 to 23, 2011. Instead of simply looking at the timelines before, during, and after the event, we differentiated the beginning of the time before, usually used for setting up things, and the very last part of the time before ("last minute"), when usually the latest news are announced and advertisements are made to give a final push in attracting attendees. The resulting phases were:

1. Setup phase (May 10-27): setup of social media accounts, website, email list, etc;



XSL•FO
RenderX

2. Active promotion phase (June 1-November 2): community growing phase before event;

3. Last-minute promotion phase (November 13-17): announcing latest information about the event a few days before;

4. Actual event phase (November 21-23): activity during the conference; and

5. Postevent phase (November 24-December 5): community behavior after the event.

6. For each phase, we measured the activity on each channel and aimed to determine its impact on the community outreach. Moreover, we measured, but we also actively generated data on those channels. This is different compared with most prior research; essentially, we let the conference become a real-world laboratory where we not only analyzed the data, but also performed research on the response to our actions. By having different phases, we could also look at which channel performed better in which phase. Naturally, the different spaces of time around an event (before, during, after a conference) will have different activities involved (eg, announcing a conference after the event has passed is quite useless). Based on the five phases and the measurements taken during these phases, we aimed to determine which channel was best suited for which phase.

### *Procedure*

Firstly, we set up the media channels as listed in Multimedia Appendix 2. The Facebook page was created to promote the event conference (eg, place and date of the conference, submission deadlines) and to provide information about the conference such as changing dates, announcing keynotes, or links to subpages of the website later on.

The Twitter account @eHealthConf was then linked to the Facebook page so that messages posted on the Facebook page were automatically posted to the Twitter account, including a link to Facebook if the message was longer than 140 characters (the limit of tweets on Twitter). To increase the number of followers of @eHealthConf, Twitter accounts of similar events known to the organizers were identified and followed along with relevant eHealth organizations, research groups, and researchers. Twitter users who followed those accounts were then also identified and followed (similar to snowball or chain sampling).

The official hashtag of the conference of #eHealth2011 was decided on and publicized via Twitter, Facebook, the conference website, and at the event itself during the welcome session. Our Liveblog system also included all tweets and retweets of the @eHealthConf account and tweets using the hashtag #eHealth2011. For the poster session at eHealth 2011, we setup a Flickr gallery where the poster presenters could upload their posters for public viewing, which was promoted during the poster session itself, on Twitter and Facebook, and also on the conference website.

All channels were linked to from the conference website and verbally promoted during the introductory session at the conference itself.

### *Measurements*

For each media channel, we defined a set of measurements for the five longitudinal phases. We took the measurements (when possible) on a daily basis (ie, summarized the value of a measurable element at the end of a day).

For Twitter, we used the following measures on a daily basis:

1. Followers: number of users following the event account (measured via the Twitter email notification on new followers).

2. Followers lost: number of users who stopped following the event account (measured via the third-party service TwUnfollow [22]).

3. Retweets: number of retweets of the event account (measured via the Twitter email notifications); this pertained only to those retweets that used the official "retweet" application programming interface (API) of Twitter. Other ways of retweeting (eg, manually writing "RT...<account name> <message>") were counted with mentions.

4. Mentions: number of tweets from other users that contained the event account name (measured via the Twitter email notifications about mentions). This did not include retweets that were done via the official retweet API of Twitter (although some clients show this as a retweet in the timeline).

5. Users receiving retweets: number of users to which messages from the event account were retweeted (via the official retweet API only). The data were derived from the Twitter email notifications (which contained information such as "@XYZ retweeted to N followers..."). We took the sum of these numbers per day, whereas when retweets of the same user occurred, we counted only once and used the maximum number (due to changes in followers of that user during the day). We did not subtract duplicates here (eg, users that received the same retweet from two or more followers of our account).

To analyze the outreach of the Facebook page, we took three measures because they are provided by the weekly Facebook status update (via email notification):

1. Likes: the number of Facebook users that liked our Facebook page;

2. Posts: the number of posts or comments on the page's wall, either made by ourselves (the page) or by others; and

3. Visits: the number of visits of the Facebook page.

On Flickr, the number of page views was the relevant measure. This could be on individual posters (or photos) or for the gallery front page as a whole.

For the Liveblog, we measured the number of online users that were connected to the service at a given time. We differentiated between the highest number of participants at any one time (for calculating maximum) and the number of total online users on a particular day. This measurement was only taken during the event because the Liveblog was only available in this phase.

The measure for the email lists was the number of email recipients. This was slightly different from the number of registered email addresses in the list because some emails could



XSL•FO
RenderX

be bounced due to various reasons (eg, address not valid or someone had set up a notification of absence).

The typical measurement of a website is the number of (daily) page visits. In addition, geographical information about the visitors could be of interest.

Finally, for the attendees, we counted the number of persons who were physically present at the conference.

### *Outreach Score*

Based on the continuously taken measurements, we calculated the outreach of each media channel for each longitudinal phase. The idea was to compare the different media channels and their outreach performance to identify the best-suited channel (or channel mix) for each phase. Previous work into promotion of scientific online content using various channels by de Quincey et al [23] was a step in the right direction, although not linked to a physical event. Therefore, we calculated three values: maximum outreach, mean outreach, and total outreach for each phase. Maximum outreach shows for each channel the maximum number of users we could reach on a single day during a particular phase. This did not necessarily mean we actually reached them because they could have missed or discarded the message. However, it was an indicator of the maximum size of a virtual community. Mean outreach for each channel was the arithmetic mean of the number of users that we reached daily during a particular phase. Again, this did not necessarily mean they actually read a message or were actively involved. However, it was an indicator of the community growth when we looked at this measure over time. Total outreach for each channel summarized the number of users that we reached in a time period of a particular phase. Although we cannot completely rule out duplicates (eg, access to the website on two different days could have originated from the same or from different users), the total outreach was an indicator how many users could be reached in total during a given phase.

We chose these outreach scores because we could not measure the exact numbers due to overlaps. Although for certain media (eg, Twitter) it was possible to rule out overlaps by using intensive data crawlers over time (capturing and analyzing the links of followers and subtracting duplicate users), we did not use this in the first place. Moreover, for some media channels, it was more or less impossible to rule out duplicate users (eg, page views on a website from the same IP address). However, we attempted to cleanse the data to reduce the potential influence of duplicates (eg, we did not summarize the number of attendees for all days of the event when they were obviously the same because we know from the registration list).

Table 1 shows the resulting outreach scores of our framework for all channels and phases. Note that all outreach scores were defined within a phase and did not include the data of the other phases. Some scores contained adjustments to reduce effects of duplication. For example, for the total outreach of Twitter during a phase, we did not summarize all followers because this would most likely include too many duplicates. Instead, we took the number of followers at the end of a phase, added the sum of the followers that we lost during this phase, and added the maximum of users receiving retweets for this phase. The latter (adding the maximum instead of the sum) is an adjustment we made because we did not know the number of followers lost of those users receiving retweets. Where we could clearly identify the users (email list and attendees of the conference), we counted the real persons as the total outreach.

**Table 1.** Outreach scores of our framework.

| Channel | Maximum outreach | Mean outreach | Total outreach |
|---|---|---|---|
| Twitter | Maximum(followers + users receiving retweets) | SUM(followers + users receiving retweets) / COUNT(followers + users receiving retweets) | SUM(followers at end of phase) + SUM(followers lost) + maximum(users receiving retweets) |
| Facebook | Maximum(likes + visits) | SUM(likes + visits) / COUNT(likes + visits) | SUM(likes + visits) |
| Flickr | Maximum(views) | SUM(views) / COUNT(views) | SUM(views) |
| Liveblog | Maximum(online users at the same time) | SUM(online users) / COUNT(online users) | SUM(online users) |
| Email | Maximum(recipients) | SUM(recipients) / COUNT(recipients) | COUNT(recipients) |
| Website | Maximum(visits) | SUM(visits) / COUNT(visits) | SUM(visits) |
| Attendees | Maximum(persons) | SUM(persons) / COUNT(days of event) | COUNT(persons) |

The reason we looked at three values (maximum, mean, and total) was that they showed a different view of community growth and interconnectivity. For example, a maximum may be very high during a phase, but this could be the result of only a single action. The mean, however, could show the density of interactions during a time period. The total shows the effectiveness over the whole period of a phase.

The essential idea of our framework was to take the outreach scores as previously defined and evaluate them for each phase. This meant these figures were repeated for each of the five phases to receive the overall view. For better comparison, the length of each phase needed to be normalized.



XSL•FO
RenderX

## Results

Table 2 details the outreach scores during each phase and compares the different media channels.

**Table 2.** Comparison of outreach results for each phase.

| Phase | Setup | | | Active promotion | | | Last-minute promotion | | | Actual event | | | Postevent | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Max | Mean (SD) | Total | Max | Mean (SD) | Total | Max | Mean (SD) | Total | Max | Mean (SD) | Total | Max | Mean (SD) | Total |
| **Channel** | | | | | | | | | | | | | | | |
| Twitter | 19 | 13 (4) | 20 | 2343 | 555 (339) | 3100 | 1011 | 721 (155) | 1023 | 1562 | 1236 (434) | 1569 | 665 | 605 (19) | 682 |
| Facebook | — | — | — | 105 | 75 (16) | 1129 | 89 | 89 | 89 | 181 | 181 | 181 | 112 | 112 | 112 |
| Flickr | — | — | — | — | — | — | — | — | — | — | — | 520 | 833 | 165 | 1489 |
| Liveblog | — | — | — | — | — | — | — | — | — | 20 | 139 (50) | 416 | — | — | — |
| Email | 1989 | 1431 | 1989 | 2046 | 2018 | 2046 | 2047 | 2047 | 2047 | 2047 | 2047 | 2047 | — | — | — |
| Website | 72 | 32 (20) | 870 | 164 | 45 (25) | 7342 | 181 | 102 (44.9) | 817 | 307 | 194 (109) | 583 | 81 | 46 (21) | 555 |
| **Traditional impact measure** | | | | | | | | | | | | | | | |
| Attendees | — | — | — | — | — | — | — | — | — | 74 | 74 | 74 | — | — | — |
| Proceedings | — | — | — | — | — | — | — | — | — | — | — | — | — | — | 4130 |
| Sum | 2080 | 1476 | 2879 | 4747 | 2693 | 13617 | 3328 | 2959 | 3976 | 4191 | 3869 | 5390 | 1691 | 929 | 6968 |

The total outreach during the event was 5390. The maximum outreach on a single day during the event was over 4191 because this did not yet include the traditional outreach of proceedings. Note that these calculations could be even higher, in particular in the other phases, because some measurements were not or could not be taken. Also, the number of proceedings (23 sold books and 4107 downloaded online articles) occurred later than our defined postevent phase, but for completeness, we added them to the total outreach of the last phase. Finally, the resulting numbers did not eliminate duplicates. For example, a physical attendee could visit the website, retweet a message from the conference Twitter account, and post something in the Liveblog while visiting the Facebook page. Hence, this is an upper bound of the outreach.

Comparing the outreach results of the different channels over the five phases, we can identify certain differences. Some channels seem to be more effective in certain phases than in others. Figures 1 to 3 show comparative diagrams for the outreach scores maximum outreach, mean outreach, and total outreach, respectively. Total outreach was normalized to the length of each phase in days.

In the first phase (setup), we had very low outreach scores in most cases because the channels had just been set up. The numbers of page visits, Twitter followers, etc, were not expected to be as large right from the beginning. One exception was the email list, which was set up very quickly (based on existing lists of recipients from previous conferences) so that a first call-for-papers could be sent out to a large number of people early. This is the traditional way of announcing a conference, in companionship with establishing a conference website.

For the second phase (active promotion), however, Twitter and email, in particular, had much more outreach than the other channels in terms of maximum and mean. If we look at the total outreach instead (see Figure 3), the traditional website has accumulated the most outreach over the period of this phase. Interestingly, this is contrary to the much higher peaks for email and Twitter in maximum and mean outreach. An explanation for this can be that Twitter is a more dynamic medium with respect to retweets and mentions, whereas website visits are rather a “static” but continuously performing outreach.

The third phase (last-minute promotion) differed a little from the previous one, although it was also related to promotional activities. Email and Twitter were still the most dominant channels in maximum and mean outreach. For the normalized total, however, Twitter overtook the website. This could be explained with the quickly increasing number of followers in this phase.

In the fourth phase (actual event), the previously “less important” channels gained more significance. The Liveblog and the attendees were present only in this phase; therefore, they had their (only) peaks here. In addition, the maximum and mean outreach of Facebook and the website also had their highest peaks here. However, email and Twitter were still the best performing channels, this time also for the (normalized) total outreach. Although one would expect those channels that offered highly dynamic interactions (eg, mentions/retweets in Twitter, likes/posts/comments in Facebook) to be the ones that outperformed all other channels, we saw that this was not true for Facebook in our case study. Liveblog, Flickr, and the website were better than Facebook in the normalized total outreach. The



XSL•FO
RenderX

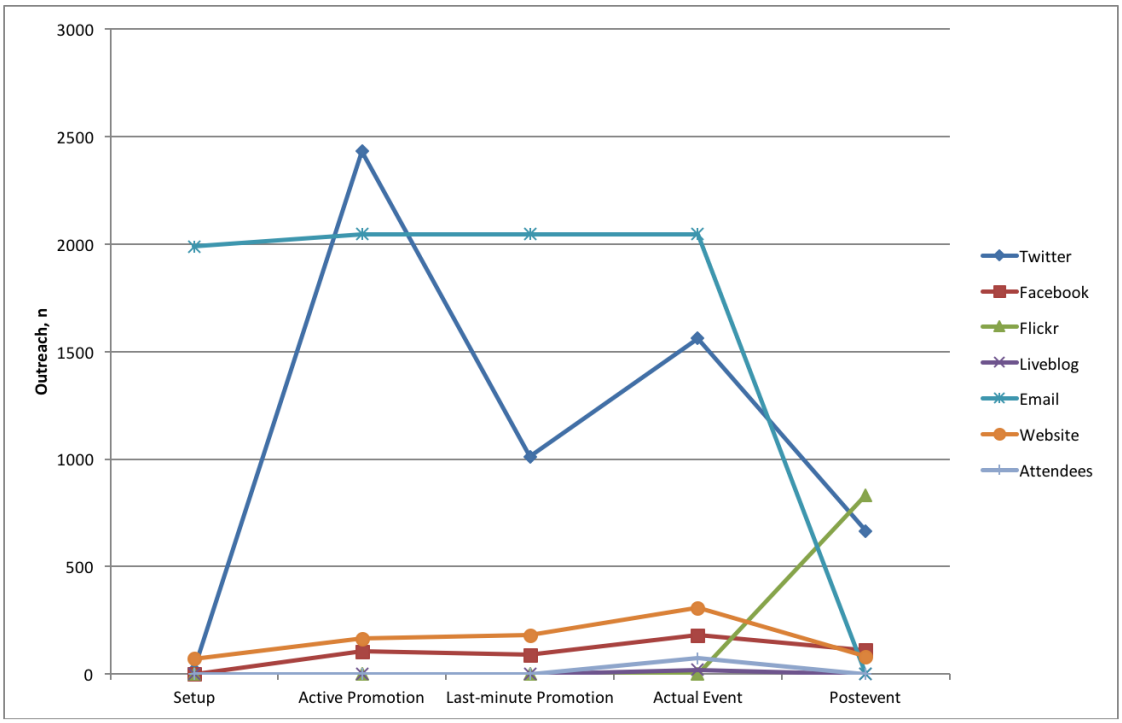
website channel was even better than Facebook in maximum outreach.

The last phase (postevent) showed a very different result. The outreach of Twitter declined (and email because no emails were sent after the event, of course). In particular, Flickr had a high outreach in all three categories (maximum, mean, and total). But, because this phase was kind of a wind down phase, it was clear that channels with a more archive-like character were more effective in this phase. This was particularly true for proceedings (which are usually read by an increasing number of people after the conference) and Flickr (as a new medium to show the conference posters to a wider community).

Twitter was far more effective than Facebook as a social media channel for a scientific conference such as in our case study. Twitter and email were the most effective channels during all phases up to the actual event. During the event, channels such as the website (eg, showing information about the program) could be enhanced with media channels that allow active participation (eg, Twitter, Facebook, and Liveblog). Interestingly, our case study showed that Facebook had less relevance, whereas the Liveblog seemed to be a good addition to support active discussion and allow people to remotely participate at the conference. Using our novel methods, the Liveblog engaged 5.6 times more "virtual participants" than those physically attending. After the event, traditional media such as proceedings (online and offline) can be enhanced with special-purpose social media such as Flickr to increase the outreach of presentations. In particular, Flickr exposure of posters gave access to seven times more users during the conference and overall, including the postconference phase, 20 times more than those who would have seen them physically. Overall, phase 2 saw the highest total (it was also longest phase), but the highest mean outreach was during the conference itself (3869 in phase 4 driven predominately by Twitter).

Although the results are only from one event (and with a relatively focused target group), they are useful as recommendations to structure and plan media channels for other (scientific) conferences. Similar to the body of socio-patterns research, the generalizability of results from various experiments with real-world participants is a challenge [24]. However, the setting of our case study around a conference makes it a standard scenario. The framework itself is generalizable and could be adapted to other events to include other media channels.

**Figure 1.** Comparison of maximum outreach of all channels.






XSL•FO
RenderX

**Figure 2.** Comparison of mean outreach of all channels.

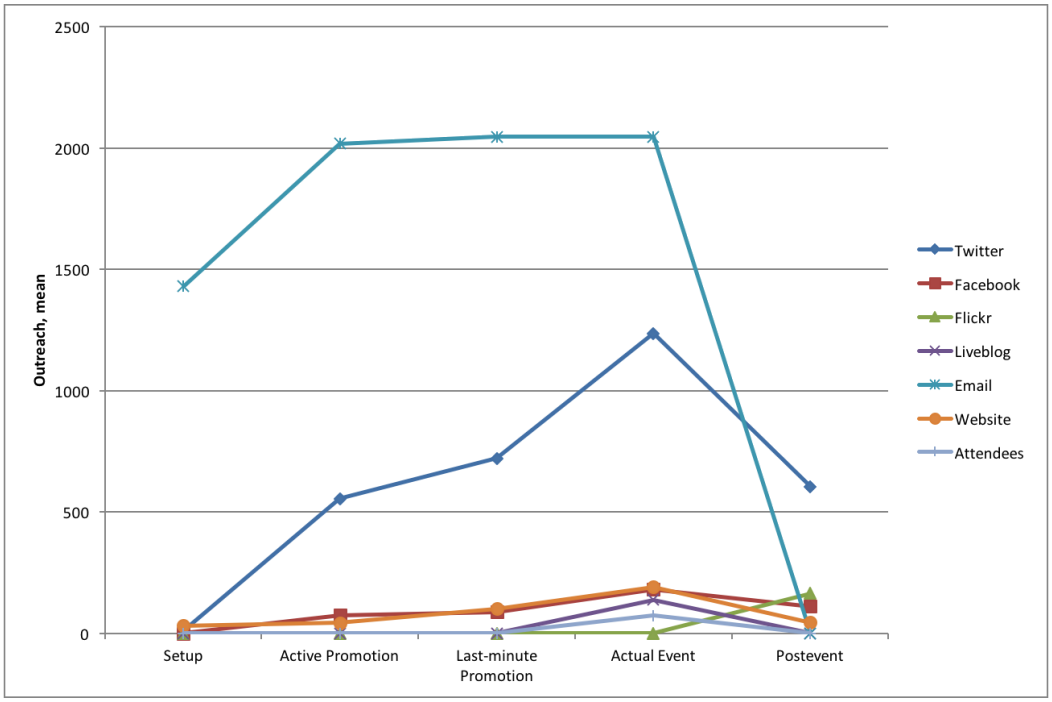
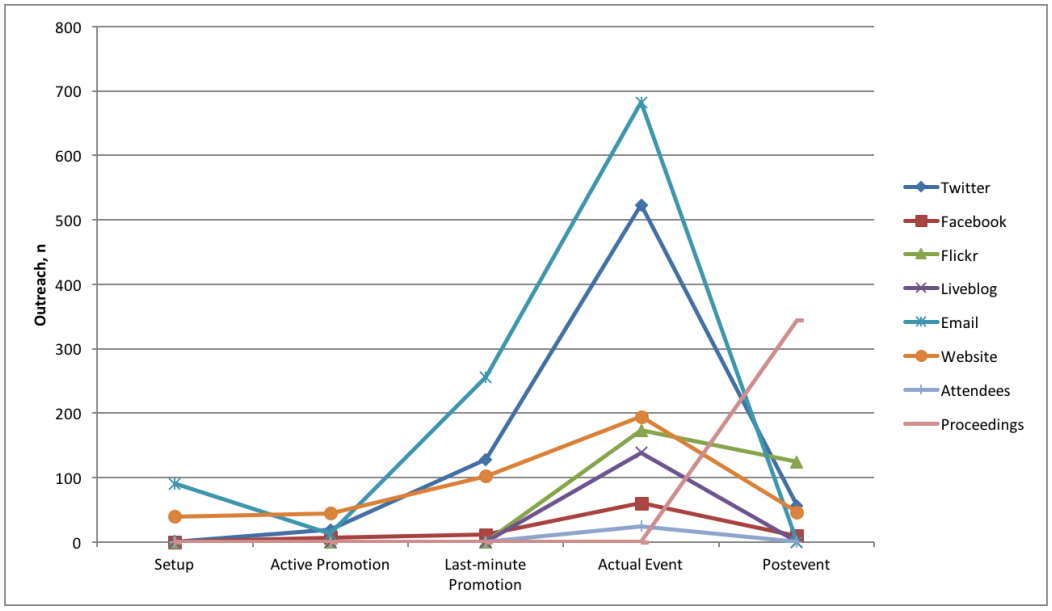



**Figure 3.** Comparison of total outreach of all channels (normalized to length of each phase).




## Analysis of Data per Channel

In this section, we present and analyze the detailed data of the different media channels we used to promote the real event. In particular, we look in more detail at the data and results on a per-channel basis.

### *Twitter*

#### Followers

Figure 4 shows the number of followers measured over time. The data were taken from email notifications from Twitter about new followers subtracted by the people who unfollowed as reported by the service "TwUnfollow" [22].

We observed that the five phases of the promotion timeline could be matched to five data periods in the graph. The first period matches with the setup phase, which had unsurprisingly low followers because the Twitter account was new and known only to the organizers themselves at this stage. The second period (active promotion phase) started with a rapid increase in the number of followers within a few days. This rapid growth matched with the time (approximately a week) when we started to follow other people (up to a maximum of 2000, a limit set by Twitter at that time to avoid spam accounts). This rapid growth was followed by moderate but continuous growth for the rest of the second phase, the active promotion phase. In the third period (last-minute promotion phase), we sent promotional and announcement tweets about the program and invited speakers. During the conference, there was a smaller increase,



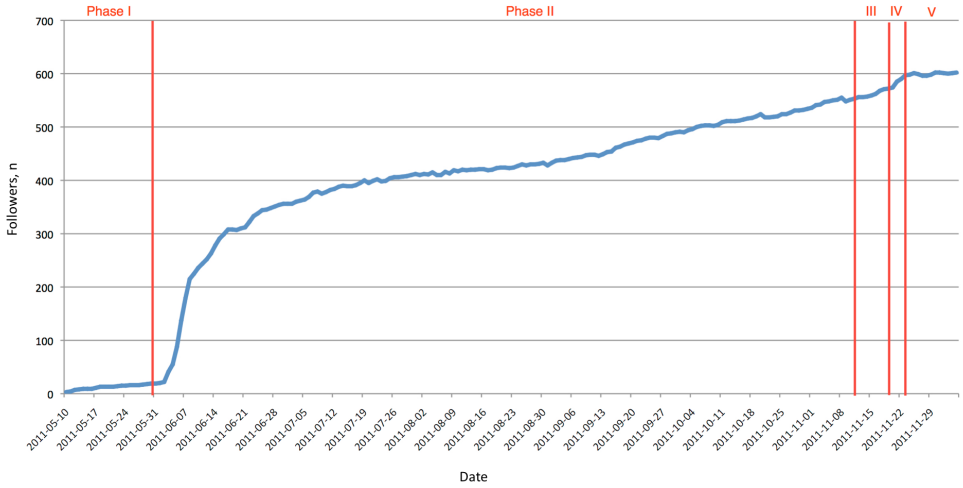
XSL•FO
RenderX

probably resulting from an increased number of tweets in the actual event phase. Finally, after the conference, the number of followers remained more or less constant.

During the short period in the beginning of the active promotion phase when we increased the users that we followed (up to 2000), a number of people followed us back immediately. However, most users did not follow us back. For example, on June 6, 2011, we followed 1579 users and 1434 did not follow us back. As our number of followers increased during the third period, we had a fairly constant ratio of followers versus following; at the end of this period (at the time of the conference), the number of users not following back was 1587, whereas we followed 1998 users. We measured the ratio over two months (September and October 2011) and during the time of the conference, but it kept almost constant. Therefore, we had a fairly constant number of people following us back (approximately 400). Some immediately followed back when we followed them, so we can only speculate about their interest in our account. We can assume some of them only followed because they were followed.

In addition, our final number of followers (more than 600) meant that we could attract approximately 200 users to follow our account without following them. We can assume they were directly interested in our account (ie, the conference and the tweets about it).

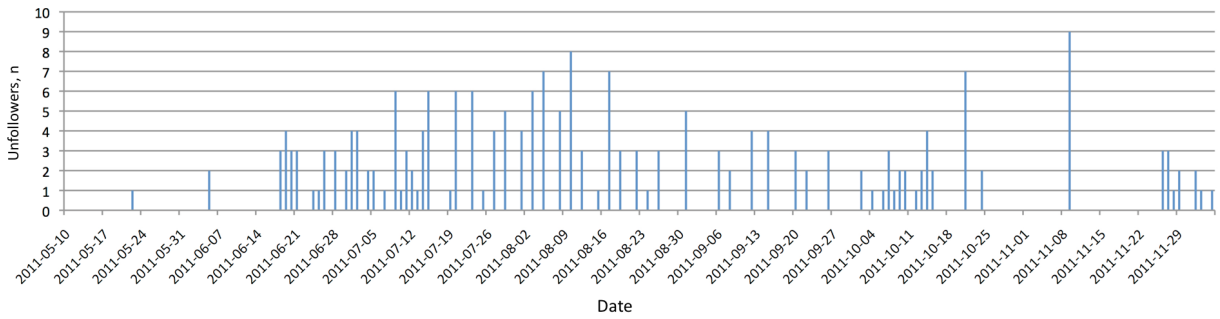
**Figure 4.** Twitter: number of followers through the different phases. Phase I: set up; phase II: active promotion; phase III: last-minute promotion; phase IV: actual event; phase V: postevent.




## Followers Lost

The number of followers we lost, as reported by the service TwUnfollow, was at a relatively low rate throughout the overall time period (see Figure 5). Although the timeline included a few peaks, as Figure 5 shows, we could not accurately assign the "unfollowers" to specific dates. The TwUnfollow service sometimes aggregated the followers lost for a few days. We assigned these numbers to the day reported by TwUnfollow. The website of TwUnfollow itself stated that due to high load "it may take up to 48 hours until unfollows appear in your history." Hence, these reports are only an approximation of specific days. Unfortunately, Twitter does not provide a comprehensive interface to analyze unfollowers.

**Figure 5.** Twitter: number of followers lost.




## Tweets and Retweets

Figure 6 shows the number of tweets that we sent through the conference Twitter account. Our tweet activity was a result of the different phases of promotion as described previously. There were three major periods of activities in tweets: (1) in June 2011, when we made initial announcements of the conference (eg, posting the call-for-papers); (2) from August to October 2011, when we announced deadline extensions and reminders



XSL•FO
RenderX

to register; and (3) November 21 to 23, 2011, during the conference itself.

The third period had the highest volume of activity because information was posted about ongoing talks and other information during the conference. The first and second periods belonged to the active promotion phase and showed that we had more activity in the first quarter and second half of this phase. However, as the number of followers before showed, there was a steady increase, even in times when we had low activity in tweets.

In addition to the tweets, we also analyzed the corresponding retweets (see Figure 7). As expected, there was a peak of received retweets during the conference (November 21-23, 2011). However, we also had a number of higher peaks before which matched the three periods of our tweet activity (of course, there would no retweets by other users to be expected if we have no tweets).

**Figure 6.** Twitter: number of tweets. The three major periods of activity (in red boxes) correspond to when initial announcements about conference were made, when deadline extensions and reminders to register were made, and during the conference itself, respectively.

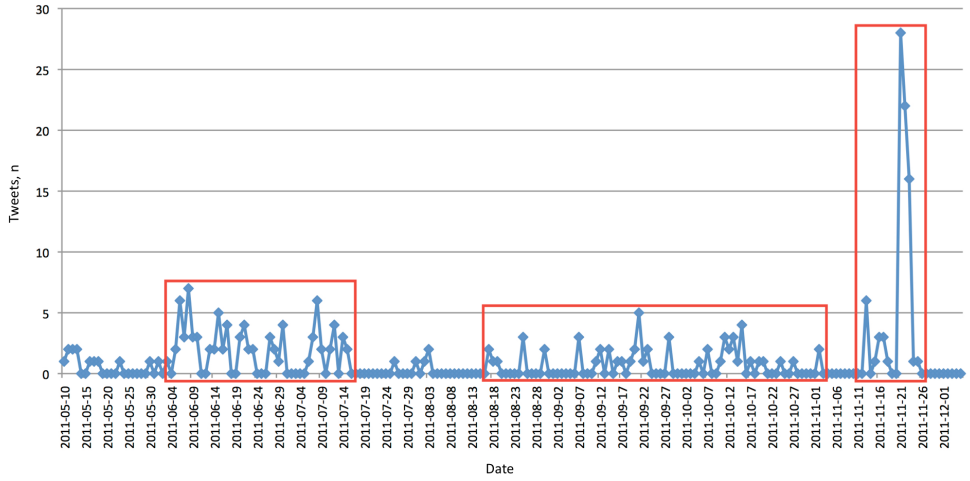
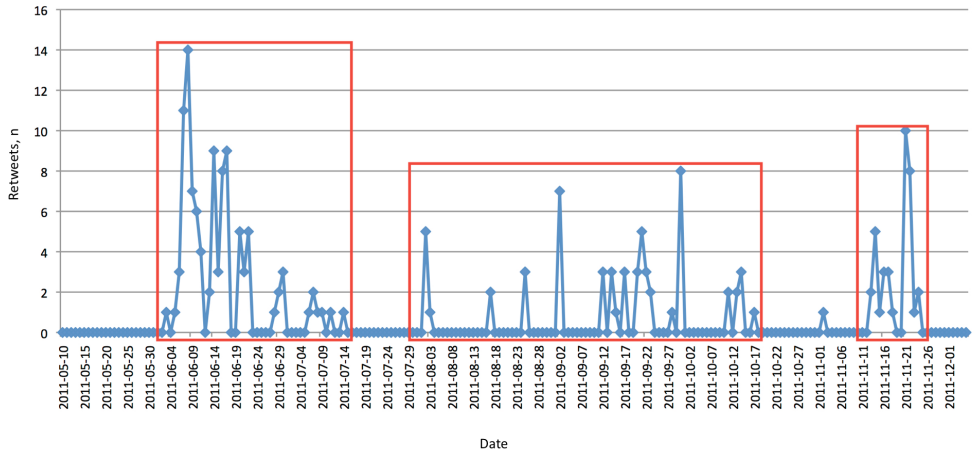



**Figure 7.** Twitter: number of retweets. The three major periods of activity (in red boxes) correspond to peaks in tweet activity.



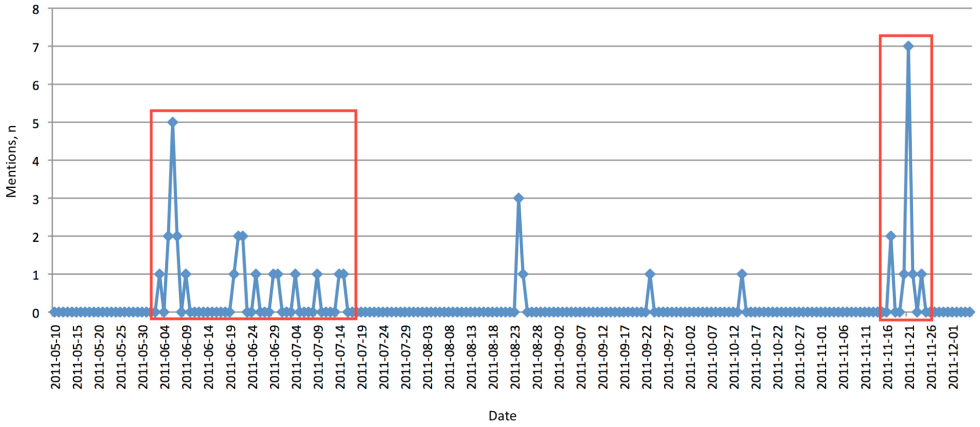

## Mentions

The number of mentions over time was also observed (see Figure 8). As pointed out by Cha et al [1], mentions is a measure for the value of a name. Because our conference Twitter account @eHealthConf did not have a long history, we did not expect too many mentions. The maximum value was indeed seven mentions on a single day (during the conference) and a few mentions over the rest of the time. Nonetheless, we observed an association between the different activities we made. There were two “dense” groups of mentions, one in the beginning and one during the conference. The former was primarily related to our activity of gaining followers by simply following many others. The latter group was unsurprisingly related to the real event of the conference itself.



XSL•FO
RenderX

**Figure 8.** Twitter: number of mentions. The two major periods of activity (in red boxes) correspond to when we were following many others and during the conference, respectively.




## Outreach

To analyze the outreach, we looked at the number of followers and added the users receiving retweets from them. For each retweet that someone made to a tweet of our Twitter account, we received an email notification from Twitter stating the number of users that received the retweet.

From Figure 9, we can observe that there are a number of high peaks. They resulted from retweets of users who had a high number of followers themselves. This meant a message from our @eHealthConf account had reached not only our followers directly, but also the followers of the user who retweeted the message. This resulted in a short-term outreach of more than 2000 users (eg, one user had more than 2400 followers and retweeted one of our messages in the early phases).

We also calculated the mean outreach over time. Figure 10 shows the mean outreach for the sum of our followers plus users receiving retweets. The mean outreach at a certain point in time included all other outreach values before (ie, we always calculated the arithmetic mean from day 1 to the current day).

From Figure 10, we can clearly observe the setup phase where there was only insignificant outreach. Then, once the active promotion phase started, there was first a sharp increase in the mean outreach, which later had slower growth. There was another small increase again in the last-minute promotion phase and during the actual event.

Based on the preceding numbers, we calculated the maximum and mean outreach of our Twitter account within the five different phases (see Table 2). If we compare the maximum and mean values in the different phases, we can make two obvious observations: (1) mean outreach was always higher in one phase than the previous, except for the last (postevent phase), which matched the continuous growth of followers and (2) maximum outreach had its highest value in the early promotion phase and another high value during the event (the former resulted from the retweeting of a single message by a user with a high number of followers and the latter was a combination of the increased number of followers for our own account and retweets by users with high number of followers).

In addition, we observed that in the last-minute promotion phase we had a lower maximum outreach (n=1011) compared to the earlier active promotion phase (n=2432) or the actual event phase (n=1562). However, the mean outreach was still growing in the last-minute promotion phase. Therefore, although the maximum outreach was lower, the increased mean outreach meant there was dense activity within the virtual community.

During the actual event, we had high values, both in maximum and mean outreach. Although the maximum outreach during the event (n=1562) was lower than the maximum outreach of the early promotion phase (n=2432), the mean outreach was high (mean 1236, SD 434). This resulted from dense activity during the conference in terms of tweeting and retweeting. In particular, it showed that the mean outreach could be higher when several people retweeted a message to only a few or moderate number of followers than the outreach induced by a single retweet of one user to a higher number of followers.



XSL•FO
RenderX

**Figure 9.** Twitter: outreach (followers plus users receiving retweets).

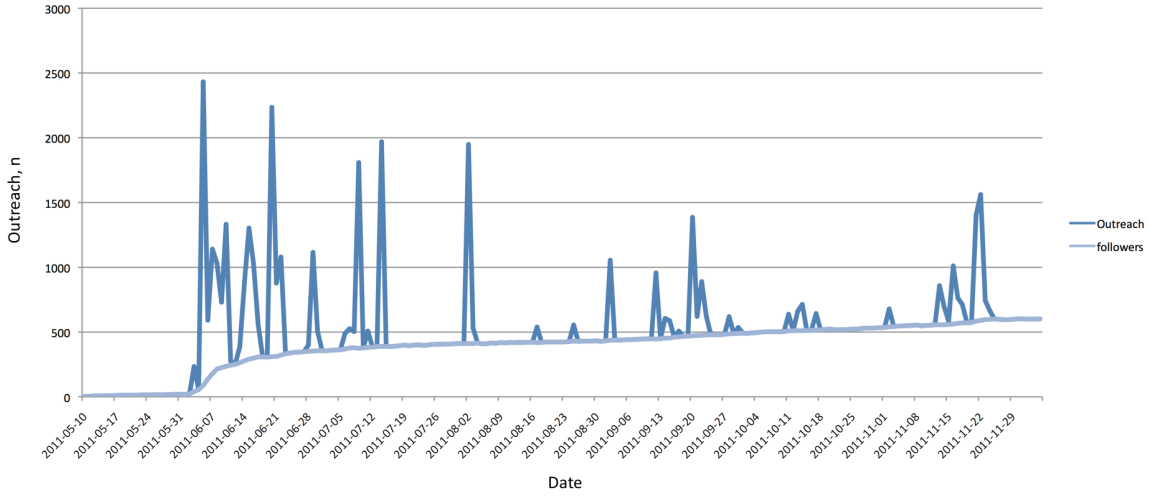
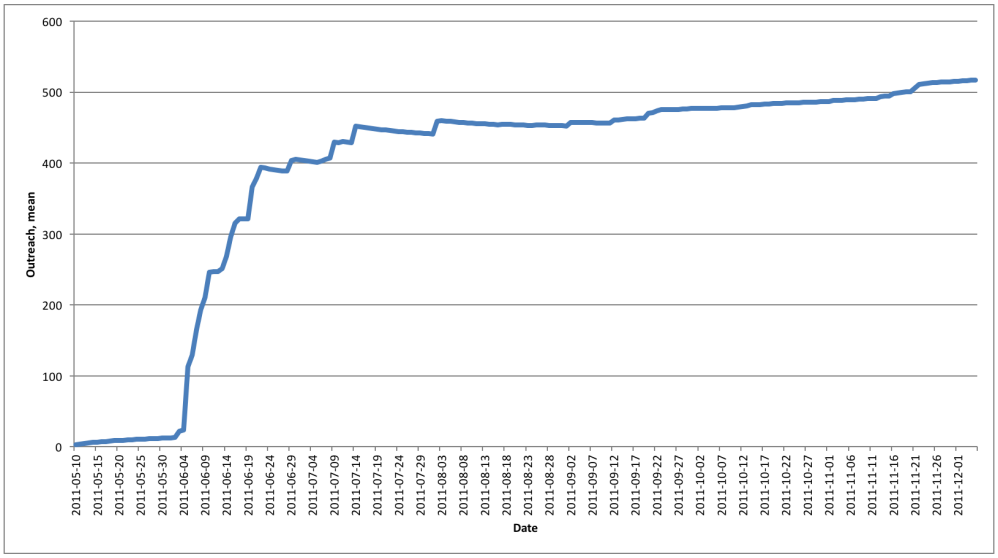



**Figure 10.** Twitter: mean outreach (arithmetic mean over time).




### *Facebook*

For each of the three measures related to our Facebook page (likes, posts, and visits), we used the data on a weekly basis because they were provided by the automatic email notifications sent from Facebook. Unfortunately, this information was only gathered from the middle of July and not from the beginning of the setup phase. Nevertheless, the data showed an increase in outreach over time, with a high peak during the actual event phase similar to the outreach of our Twitter account.

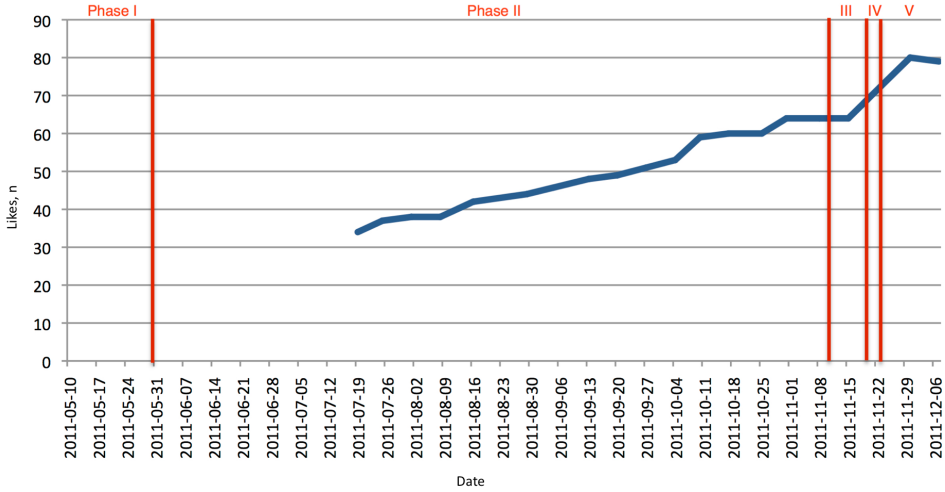
#### Likes

The number of Facebook users that “liked” our Facebook page continuously grew from only a handful (actually the Facebook user accounts of the event organizers) up to approximately 80 at the time of the conference. The growth was almost linear as Figure 11 shows.



XSL•FO
RenderX

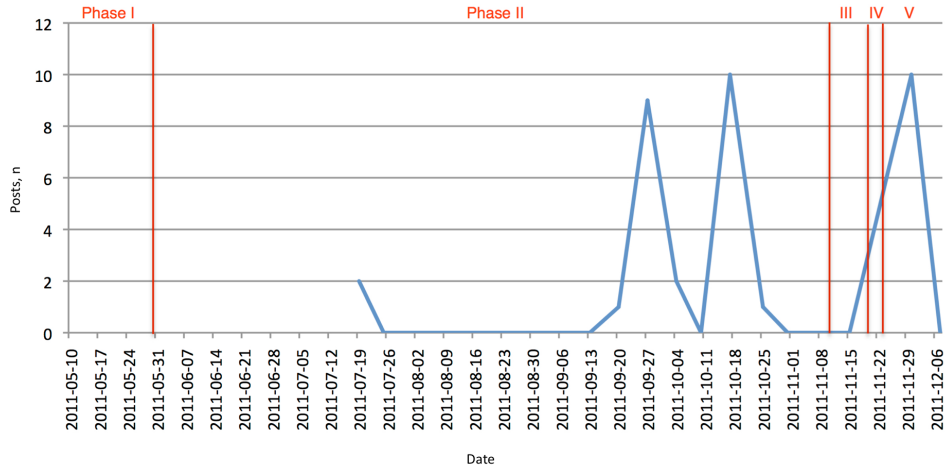
**Figure 11.** Facebook: number of likes. Note: information was only gathered from mid-July onward. Phase I: set up; phase II: active promotion; phase III: last-minute promotion; phase IV: actual event; phase V: postevent.




## Posts

We counted both posts made by us on the page and also posts and comments by other users as "posts." This measure was already included in the weekly email notifications we received from Facebook. We observed three peaks in posts/comments (see Figure 12). The first two were in September and October; this is when we posted information about the deadlines, announced invited speakers, and posted reminders about registration and the conference program. The third peak occurred around the event itself, which included information about the invited speakers and updates on the conference program.

**Figure 12.** Facebook: number of posts. Note: information was only gathered from mid-July onward. Phase I: set up; phase II: active promotion; phase III: last-minute promotion; phase IV: actual event; phase V: postevent.



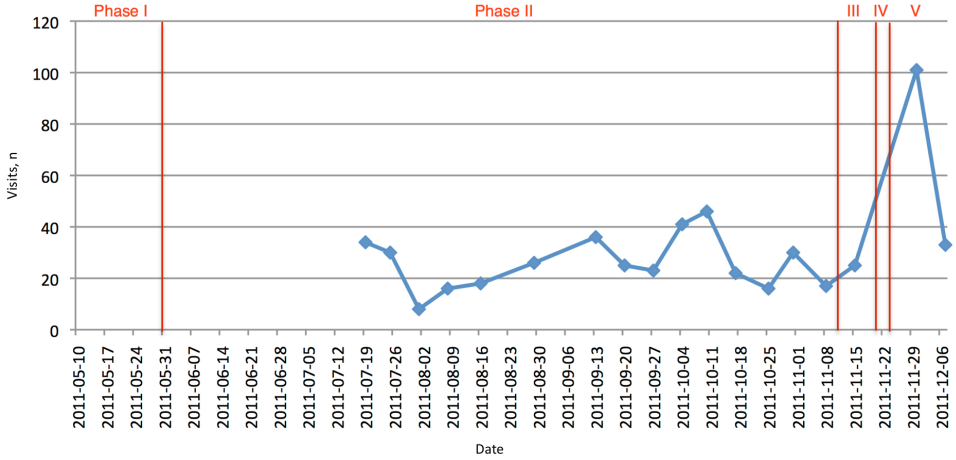

## Visits

The Facebook weekly statistical notifications also included information about the actual visits to the page. These numbers reflected the number of people who actually looked at the page (ie, by following a post that appeared in their "news" timeline, by loading the page specifically, or by following an external link such as our Facebook-Twitter link). Figure 13 shows that there were a number of smaller peaks during the promotion phase and during the time the event took place.



XSL•FO
RenderX

**Figure 13.** Facebook: number of visits. Note: information was only gathered from mid-July onward. Phase I: set up; phase II: active promotion; phase III: last-minute promotion; phase IV: actual event; phase V: postevent.




## Outreach

A measure of outreach via Facebook was the sum of the likes and the visits. These values reflected the actual readers (visits) and the potential users (likes) that we could reach in each phase. Unfortunately, we did not have a complete dataset, but only a weekly update. Hence, the outreach was determined on a weekly basis, which resulted in the same values for maximum and mean outreach for the previous three phases (see Table 2). Moreover, as the datasets started with mid-June, our outreach analysis also missed the values for the setup phase. Nevertheless, the results showed that we had increasing outreach up to the time of the actual event, with maximum and mean outreach both at 181 during the conference.

### *Website*

Information about our conference was also available on the main conference website. This was the main information site and accounts on Twitter and Facebook always contained links to the website. The website was also the only media used to submit papers and to register for the conference. We logged the daily page visits on our website during the same time period as analyzed previously (ie, from May 2011 until begin of December 2011). Figure 14 shows the daily page visits (for the entire website) during that period.

We observed a number of high peaks (more than 100 page visits per day) on the following dates:

June 22, 2011: slightly increased number of tweets on that day and the previous two days announcing membership of senior technical program committee and selective topics from call-for-papers (total: nine tweets on June 20, 21, and 22).

August 25, 2011: three tweets on that day (venue confirmed and link to page on website, registration open and link to page on website, announcement of extension of the poster/demo submission deadline).

September 22, 2011: slightly increased number of tweets on that day and the previous two days announcing confirmed keynote speakers and link to page on website.

September 29, 2011: no clear potential cause from Twitter (there were three tweets on the day before, but they were only retweets of news from other and no link to our website).

November 14-18, 2011 (just before the conference): probably people wanted to check the latest news/changes to the conference (eg, detailed program, when the conference starts, where the hotel venue was).

November 20-23, 2011 (during the conference): assume this is primarily due to the live blogging of the conference talks from an analysis of the geographic locations of the origins of these accesses (discussed subsequently).

The preceding explanations are only potential reasons for the high peaks in the website visits because we cannot make direct correlations due to missing tracking capabilities.

Using Google Analytics, we found the geographical locations of visitors to the website. During the conference, there were accesses from 44 different countries (see Figure 15), which compared favorably with the number of countries represented by the conference delegates (24 different countries) and seemed to indicate that the website had a higher outreach than the physical attendance at the conference (of course, delegates physically attending the conference would receive much more information and individual benefit than those viewing the website, so this is a measure of the geographical outreach rather than the absolute impact). There were a large number of accesses from Spain; by using Google Analytics, we saw that 79% (198/250) of these were from Malaga. It seems most likely that the majority of these were from delegates in the conference venue.



XSL•FO
RenderX

**Figure 14.** Website: number of visits. Phase I: set up; phase II: active promotion; phase III: last-minute promotion; phase IV: actual event; phase V: postevent.

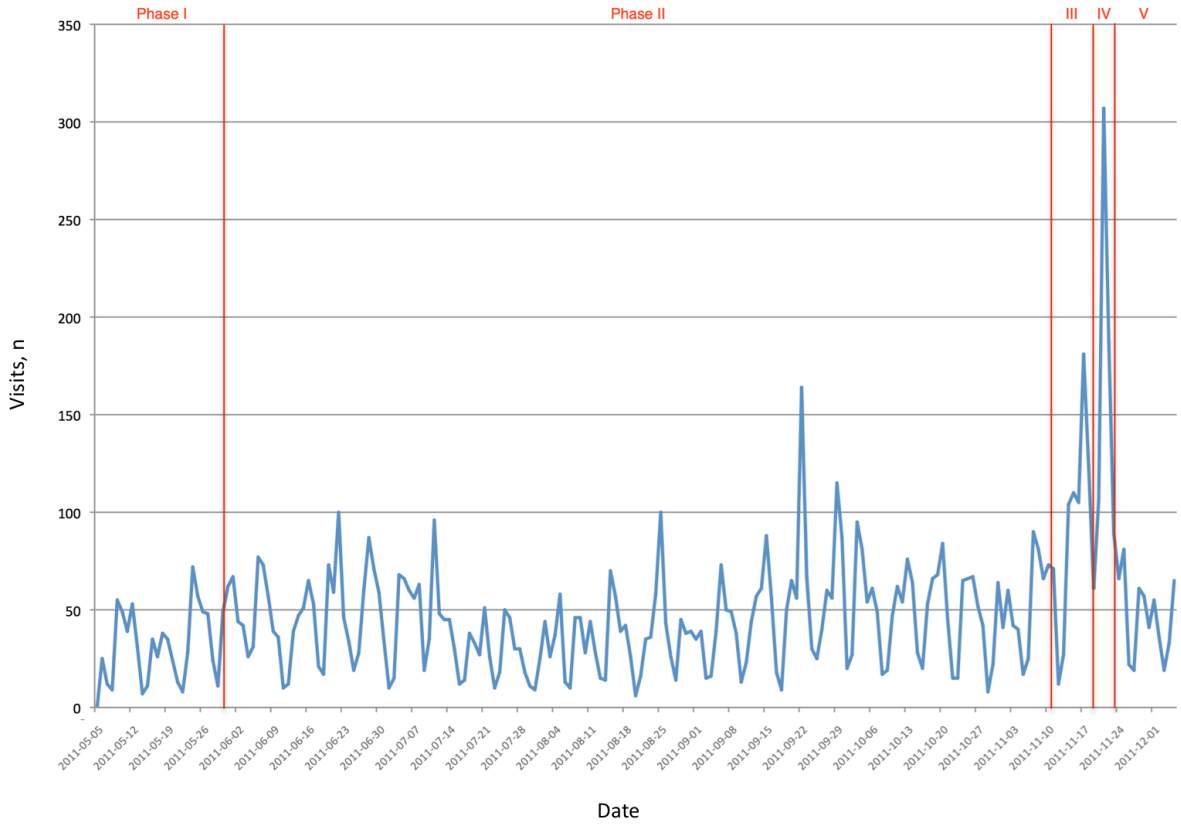
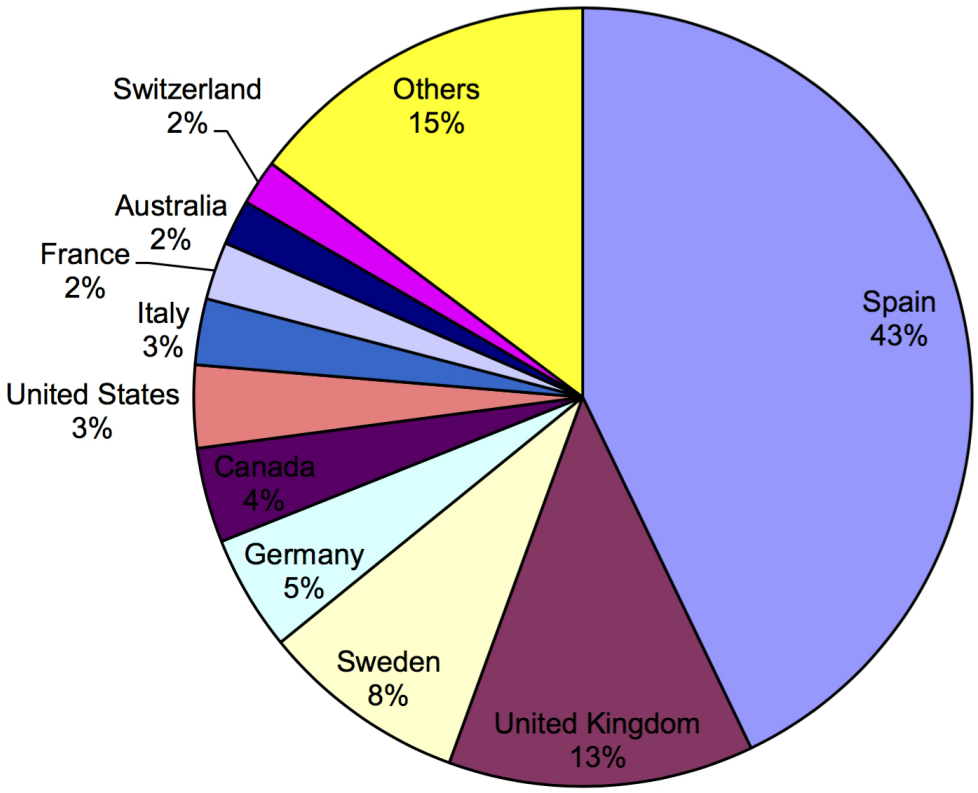



**Figure 15.** Website: visits by location.






XSL•FO
RenderX

### *Flickr*

Poster presentations at conferences are a long-standing academic staple. Their popularity has increased in the scientific community due to their ability to quickly and efficiently communicate research, and a number of guides outlining what makes a good poster and a good poster session have been proposed [25,26]. Their general success in disseminating research activity has been widely reported (eg, [27]) but, unlike a conference paper, their impact within the academic community is time limited because they tend not to have a life outside of the conference or after the conference has finished. Although abstracts are often published to accompany the poster session, a great deal of information that is contained within the poster is often lost. Therefore, a possible solution to this problem is to make the posters themselves available to delegates, perhaps in printed form or electronically on a USB memory stick or a CD-ROM, but again this is restricted to conference delegates, many of whom will have already had the opportunity to attend the poster session.

During the previous edition of the conference (eHealth 2010), one of the authors presented a poster at the conference and also uploaded it to the popular photo sharing website Flickr along with the abstract in the description. At the time of writing, this poster has now had more than 8000 views in approximately 500 days (a mean of approximately 16 views per day). In comparison with the number of attendees at the conference poster session in 2010 who saw the poster, this is a significant increase and presents a potential method for increasing the number of views of posters and, therefore, increasing the impact and outreach of the research they represent.

#### Method

Following the success of the poster described previously, it was decided by the organizing committee of eHealth 2011 that in addition to the traditional poster session at the conference, authors would be asked to participate in an online poster session. Abstracts for the posters went through the usual peer-review process and the authors of accepted abstracts were sent instructions to upload a version of their completed poster before the conference to Flickr along with the abstract in the description and add it to a public eHealth 2011 group created by the poster chair. Six of the 10 posters were successfully added to the group, but problems were reported when attempts were made to add the posters of one of the authors. Flickr had a policy regarding recently created accounts adding photos to groups (to protect against spamming) and three posters that had been uploaded to Flickr could not be added to the group. One of the key objectives of adding photos to a group was to allow the conference organizers to provide a single link to the posters on the conference websites and related promotional activities. However, using groups in this manner was not suitable, so a workaround was found in the form of galleries. Galleries on Flickr are “a way to curate up to 18 public photos or videos of your fellow members into one place” [28]; therefore, the poster chair created an eHealth 2011 gallery and added the nine uploaded posters to the gallery and it was this link that was then promoted.

#### Results

Unfortunately, Flickr only allows users to view the distribution of page views over time within a rolling 28-day period. Therefore, the results presented here are limited to total page views only. The eHealth 2011 gallery front page was set up on November 20, 2011, during the actual event phase, with nine posters and received 520 views. However, individual posters received more page views indicating referrals from other sources (eg, search engines with indexed keywords in the poster title and description). Table 3 shows the number of page views for each poster.

**Table 3.** Number of page views on Flickr for each poster.

| Poster title | Page views, n |
|---|---|
| The Guidance for Review and Approval of the U-health Care Medical Device | 51 |
| Enhancement of Sensitivity with Gathering Internet-Based Systems for Early Threat Detection Within the Global Health Security Initiative (GHSI): The EAR Project | 43 |
| Social Networks and Medical Doctors and Students | 833 |
| Epidemic Intelligence (EI) in France: Social Networking Emphasising the Process | 51 |
| A Remote Elderly Assisted Living (REAL) System | 121 |
| Representing and Accessing Scientific Knowledge About the Alzheimer’s Disease: The Semantic BiblioDem Portal | 55 |
| Review of Evaluation Processes of Web-based Systems Mining Medical Information Applied to Epidemic Intelligence | 36 |
| Reinforcing Antimicrobial Pharmacology Knowledge of Health Science Students Through a Tower Defense Video Game | 138 |
| Connect and Share: Helping Seniors with Social Isolation Use Facebook | 161 |
| Total | 1489 |

Since November 2011, during the postevent phase, the mean number of page views per poster was 165, but a single poster received the majority of the page views (n=833). Due to not having access to the referral data for the posters, it is hard to determine why this poster received more page views than the others combined. A potential explanation is that the user may have already had a following on Flickr, but this was the only upload that this user made. More likely explanations are that the author had an active Twitter account with more than 300 followers where a link to the poster was posted and also the keyword-friendly title of the poster (“Social Networks and



XSL•FO
RenderX

Medical Doctors and Students") may have drove traffic via search engine referrals.

### Conclusions

In total, the online poster session created more than 2000 page views to the gallery and posters. In comparison to high-traffic sites, this figure could be seen as insignificant, but there were approximately 30 delegates present at the eHealth poster session and the online poster presentation represented a percentage increase in views of approximately 6500%. In addition to page views, an online poster session increased the lifetime and permanency of the posters and also had the potential to promote discussion during the conference and after the conference ended (although no comments were made at the time of writing on any of the individual poster pages).

## *Liveblog*

To make the conference more exciting for virtual participants, we ran a live blogging service by a dedicated reporter (science journalist) who attended the event specifically in this capacity. The live blogging platform CoverItLive [29] was used during the conference to implement the Liveblog. This allowed the reporter to provide live coverage during the conference so that an external audience could follow the proceedings, comment, and question the participants. The reporter also acted as an online moderator and could, if it proved necessary, block unsuitable comments and spam from being published. Very few external users commented via CoverItLive. Only three comments were made in this way and, of these, only two of these were published after moderation.

Another benefit of CoverItLive was that it also aggregated Twitter messages using the conference hashtag, which was by far the most common route for participators to make comments. On analysis of CoverItLive's statistics, there were 416 readers of the live blog, although information on their locations was not available. At the end of the conference, the social media aspects of the conference were archived by allowing a replay of the Liveblog on the conference website.

**Table 4.** Summary of information gathered from the CoverItLive blog.

| Information category | | n |
|---|---|---|
| **Reader information** | | |
| | Total readers | 416 |
| | Email reminders set | 0 |
| **Published entries** | | |
| | Reporter comment | 180 |
| | Twitter comment | 375 |
| **Reader comment** | | |
| | Reader comments sent | 3 |
| | Reader comments published | 2 |
| **Media count** | | |
| | Images shown | 1 |
| | Newsflashes | 1 |
| **Google Analytic** | | |
| | Number of replays | 1507 |
| Highest number of participants noted on blog at any one time (Tuesday morning) | | 20 |

Comments made by the readers were positive and ranged from an acknowledgment of being able to access and follow the conference (eg, "Learning a lot from this; thanks guys") to more specific evaluations and questions (eg, "Ruth Hunter gave a great talk on the novel systems for behavior change...would be great to learn more about what motivates different age cohorts").

### Conclusion

The CoverItLive blog increased the reach of the conference to a wider audience with external participants logging in to make a connection with the conference output. The provision of the blog was effective in creating a wider and engaged audience, which allowed the conference to have a greater impact. The direct questioning of some of the speakers at the conference by the external participants demonstrated a physical community coming together virtually to take part in a real-time conference event. This suggests a future model for widening participation and the impact of scientific conferences.

Despite the lack of information about the location of the audience, it was clear that many people found the information provided useful as evidenced by the high number of views that could be seen by examining the Google Analytics information.

## *Email Lists*

To further promote the event, we used existing Yahoo! email lists with a combined total of approximately 300 users to send the call-for-papers and other conference announcements. In addition, we used our own list of nearly 2000 email addresses of participants at past conferences and other events. The numbers of email addresses on the latter list at the beginning



XSL•FO
RenderX

and end of the various phases are shown in Table 5. Most of the email gathering activity occurred during the setup phase. Furthermore, we estimated that approximately 5% to 10% of the emails that were sent bounced. Note that the maximum, mean, and total outreach results in Table 2 include the numbers of all email lists combined.

**Table 5.** Number of email recipients in our own list.

| Phase | Email addresses at phase start, n | Email addresses at phase end, n |
|---|---|---|
| Setup | 652 | 1828 |
| Active promotion | 1828 | 1888 |
| Last-minute promotion | 1888 | 1889 |
| Actual event | 1889 | 1889 |
| Postevent | 1889 | 1889 |

## Discussion

Our results show that the outreach of a scientific conference can be much higher than measured by traditional impact measures (ie, conference attendees and published proceedings). Despite this physical community during and after the conference, we also took online channels and social media into account. Although many scientific conferences today already use email lists and websites to promote their event, our case study shows that additional outreach can be achieved through social media. Based on the presented data, we can observe that the impact of the different types of channels varies:

Flickr: perhaps most effective (postevent) is the use of a photo gallery service such as Flickr to build a permanent "virtual poster session." Essentially, we were able to increase the outreach of normal poster session attendees of approximately 30 people to approximately 2000 views on the Flickr gallery.

Liveblog: the use of the hashtag on Twitter and its integration into the live blogging service operated during the conference brought in more interactivity than with physical attendees only. Thus, additional questions were raised from the virtual community that were not present at the conference and members of both the virtual and physical communities discussed the actual conference presentations online.

Twitter/Facebook: Twitter was best in terms of creating a longitudinal complex stream of information (600 followers receiving regular updates vs 80 likes on Facebook). It seems that the scientific community (at least for this conference) prefers Twitter over Facebook to be informed about and discuss conferences. The popularity of Twitter might be that it is a public medium as opposed to the closed network that Facebook supports. Moreover, at the time of the conference there was no equivalent of the hashtag on Facebook. However, we cannot make a general statement here as we only measured the data for one event. In addition, we have to say that approximately 400 of the 600 Twitter followers might be a result of "follow back" behavior (ie, following other users once they are followed by them). Nevertheless, we can assume the remaining 200 followed the account driven by their own interest.

In summary, our results show that the use of Twitter and email are most effective in terms of outreach in the phases before and during the event. However, the website accumulated the highest total in the active promotion phase and Liveblog, Facebook, and the website gained more attraction during the event. For the postevent phase, although there was still some activity on Twitter, the "natural winners" are those channels that have archiving characteristics (ie, Flickr for the poster session and the offline and online proceedings). Based on these results, it could be advantageous to focus activities on these channels according to our phases.

### Limitations

This study also has a number of limitations. Most importantly, we analyzed each channel individually on its own. We did not try to identify or analyze interconnections between the various channels. For instance, it would be interesting to see how many tweets brought users to the website and the other way round, and how many clicks to the Twitter box from the home page resulted in a retweet. There might also be duplicates in terms of the actual people behind the different media users. This is why we can only speculate about the actual outreach. A more complete analysis would try to remove duplicates and identify the links between the different channels. However, some links may not be detectable (eg, anonymous webpage visitors cannot be tracked to their potentially existing social media accounts).

In addition, for the postevent phase, we only took data from a 12-day period after the conference. Of course, increased outreach could be achieved if measured for a longer period. For instance, proceedings have actually been printed and distributed a couple of months after the event. Moreover, individual paper downloads usually occur a long time after the event. Although access to the poster gallery on Flickr continues, the paper downloads are a traditional means of measuring outreach and will most likely not influence the "new" means of outreach. However, maintaining a virtual community for an extended period after a real event might be useful and important (eg, to support the next event related to the previous one because, typically, scientific conferences happen annually).

One final limitation of this study is that factors relating to effort and cost-effectiveness have not been considered. There is clearly the potential to reduce effort in some of the activities (eg, by linking social media accounts so that a single post appears on various channels), but there were considerable costs for the authors with respect to time taken to create the accounts, formulate strategies, and then create and post content. For practical use, these factors need to be carefully identified and further work is needed in this area to gauge whether the use of



XSL•FO
RenderX

freely available channels such as Twitter are cost-effective in comparison to more static channels, such as the website, which have considerable setup costs.

### Future Work

Future work could try to find the interconnectivity in the outreach of the different media channels and the link between the virtual community growth and the real community behind it. For some channels, more intensive data crawling could help to find these analyses (eg, linking the Twitter users that received retweets to previous retweets or other involvement). Other findings might only be possible by tracking known users or asking them for their consent to reveal their online identity and use this data (eg, linking a user on Twitter to a real person who might be an attendee at the conference and a visitor of the website).

Further research could aim to generate a social network and understand the topological changes caused by such an event (eg, what is the rate of the increased density of the social network as a result of meeting in person at a conference?). A possible way to collect a richer dataset for this purpose would be to start the usage of Twitter earlier and build up a larger number of followers. For example, the USENIX Association has a Twitter account (@usenix) with more than 4100 followers (at the time of writing this paper), which has existed since November 2008. This account is reused to announce and promote various conferences organized by USENIX. An important factor here is to maintain the community for a series of events. Applying our approach to accounts and organizations such as these would allow for even larger longitudinal analyses.

Finally, future work could add live stream audio/video from the conference to the website to better engage with virtual participants. Unfortunately, this was not possible at eHealth 2011 for local logistical reasons. Other conference series use this type of media already to increase their outreach (eg, the Chaos Communication Congress and the USENIX Annual Technical Conference provide audio and video records of their events on their website). However, the effects of this outreach and its meaning compared to other media channels have not been analyzed in a longitudinal fashion yet.

### Conclusion

In this paper, we present a robust framework to define a physical and virtual community around an event and the role and effectiveness of online and social media usage in the promotion and presentation of a scientific conference.

The main approach is to establish a virtual community around the physical community of the real event; we also established five phases for event promotion (setup, active promotion, last-minute promotion, actual event, postevent) with the aim to observe the community growth behavior over the five phases around the event. In contrast to existing works, we made a comparative analysis of the media channels and a longitudinal study rather than looking at snapshots of data from a single medium. We also combined a virtual and a physical community, analyzed their growth and behavior over the five phases with respect to dissemination of scientific outputs and outreach, and we measured the outreach and engagement by two-way communication (ie, we were promoting the event and mining the data about the promotion at the same time). To illustrate our approach, we presented a case study of a real scientific event, the eHealth 2011 conference, which took place in Malaga, Spain, in November 2011. As we ran the conference, we also had a unique opportunity to develop and measure the outreach strategy of the conference with full understanding and insight into the social media strategy rather than just analyzing social media data of a random event.

The main achievement was the novel generalizable framework and we found insights into one conference outreach using our novel method. Our framework includes five phases for event promotion (setup, active promotion, last-minute promotion, actual event, postevent), defines virtual and physical communities, defines outreach and impact measures, and provides guidelines to measure the outreach of separate social media channels. Results from our case study of the eHealth 2011 conference revealed that it seems advantageous to focus on different media channels in each of the five phases: A mix of Twitter, email, and a website can be recommended to achieve the highest outreach before the conference, and these channels can be extended with Facebook and a Liveblog during the event, whereas the best channels after the event were Twitter and (for the long run) Flickr and proceedings. This is a cornerstone of research into a more robust understanding and analysis of social media promotion strategies for conference organizers who wish to apply our framework.

### Acknowledgments

We like to thank David Fowler for his contributions to the organization of the eHealth 2011 conference as well as to the collection and analysis of the data for this paper. We also like to thank Gozde Zorlu for producing the content of the Liveblog.

### Conflicts of Interest

None declared.

### Multimedia Appendix 1

Comparison of traditional and new impact measures.

[PDF File (Adobe PDF File), 24KB - jmir_v18i7e191_app1.pdf ]



XSL•FO
RenderX

### Multimedia Appendix 2

Media channels of the eHealth 2011 conference.

[PDF File (Adobe PDF File), 23KB - jmir_v18i7e191_app2.pdf ]

### Multimedia Appendix 3

Promotion timeline of the eHealth 2011 conference.

[PDF File (Adobe PDF File), 30KB - jmir_v18i7e191_app3.pdf ]

## References


1. Cha M, Haddadi H, Benevenuto F, Gummadi K. Measuring user influence in Twitter: the million follower fallacy. Palo Alto, CA: AAAI; 2010 Presented at: 4th International AAAI Conference on Weblogs and Social Media; May 23-26, 2010; Washington, DC.
2. Romero D, Galuba W, Asur S, Huberman BA. Influence and passivity in social media. In: Proceedings of the 20th International Conference Companion on World Wide Web. 2011 Presented at: WWW'11: 20th International Conference Companion on World Wide Web; Mar 28-Apr 1, 2011; Bangalore p. 113-114. [doi: 10.1145/1963192.1963250]
3. Russo P, Nov O. Photo tagging over time: A longitudinal study of the role of attention, network density, and motivations. In: Proceedings of 4th International AAAI Conference on Weblogs and Social Media. Palo Alto, CA: AAAI; 2010 Presented at: 4th International AAAI Conference on Weblogs and Social Media; May 23-26, 2010; Washington, DC.
4. Cosley D, Lan X. Sequential influence models in social networks. In: Proceedings of 4th International AAAI Conference on Weblogs and Social Media. Palo Alto, CA: AAAI; 2010 Presented at: 4th International AAAI Conference on Weblogs and Social Media; May 23-26, 2010; Washington, DC.
5. PAWS Lab. iTunes App Store. 2011. UMAP 2011: Conference Navigator URL: https://itunes.apple.com/us/app/umap-2011/id446590973?mt=8 [accessed 2016-02-10] [WebCite Cache ID 6fC7rzFx1]
6. Khrouf H, Atemezing G, Rizzo G, Troncy R, Steiner T. Aggregating social media for enhancing conference experience. In: Workshop on Real-Time Analysis and Mining of Social Streams. 2012 Presented at: Sixth International AAAI Conference on Weblogs and Social Media; June 4–7, 2012; Dublin, Ireland.
7. Meeder B, Karrer B, Sayedi A, Ravi R, Borgs C, Chayes J. We know who you followed last summer: inferring social link creation times in Twitter. In: Proceedings of the 20th International Conference on World Wide Web. New York: ACM; 2011 Presented at: WWW '11; March 28-April 01, 2011; Hyderabad, India.
8. Bruns A, Woodford D, Sadkowsky T. First Monday. 2014 Apr. Towards a methodology for examining Twitter follower accession URL: http://firstmonday.org/article/view/5211/3864 [accessed 2016-02-10] [WebCite Cache ID 6fC81P91x]
9. Golbeck J. First Monday. 2007 Nov. The dynamics of web-based social networks: membership, relationships, and change URL: http://firstmonday.org/article/view/2023/1889 [accessed 2016-02-10] [WebCite Cache ID 6fC7M1uKv]
10. Tufekci Z. Who acquires friends through social media and why? 'rich get richer' versus 'seek and ye shall find. In: Proceedings of 4th International AAAI Conference on Weblogs and Social Media. Palo Alto, CA: AAAI; 2010 Presented at: 4th International AAAI Conference on Weblogs and Social Media; May 23-26, 2010; Washington, DC.
11. McKendrick DR, Cumming GP, Lee AJ. Increased use of Twitter at a medical conference: a report and a review of the educational opportunities. J Med Internet Res 2012;14(6):e176 [FREE Full text] [doi: 10.2196/jmir.2144] [Medline: 23232765]
12. Wen X, Yu-Ru L, Trattner C, Parra D. Twitter in academic conferences: usage, networking and participation over time. In: Proceedings of the 25th ACM Conference on Hypertext and Social Media. New York: ACM; 2014 Presented at: 25th ACM Conference on Hypertext and Social Media; Sep 1–4, 2014; Santiago, Chile p. 285-290 URL: http://arxiv.org/pdf/1403.7772v1.pdf [WebCite Cache ID 6iqwyAmsL]
13. Reinhardt W, Ebner M, Beham G, Costa C. How people are using twitter during conferences. In: Creativity and Innovation Competencies on the Web. 2009 Presented at: 5th EduMedia Conference 2009; May 4-5, 2009; Salzburg p. 145-156.
14. Letierce J, Passant A, Breslin JG, Decker S. Using twitter during an academic conference: The #iswc2009 use-case. In: Proceedings of the 4th International AAAI Conference on Weblogs and Social Media. Palo Alto, CA: AAAI; 2010 Presented at: 4th International AAAI Conference on Weblogs and Social Media; May 23-26, 2010; Washington, DC.
15. Letierce J, Passant A, Breslin JG, Decker S. Understanding how twitter is used to spread scientific messages. In: Proceedings of the WebSci10: Extending the Frontiers of Society On-Line. 2010 Presented at: WebSci10; April 26-27, 2010; Raleigh, NC.
16. Weller K, Dröge E, Puschmann C. Approaches for defining and measuring information flows within tweets during scientific conferences. In: Proceedings of the 1st Workshop on Making Sense of Microposts. 2011 Presented at: 1st Workshop on Making Sense of Microposts (#MSM2011); May 30, 2011; Heraklion, Crete.
17. Desai T, Shariff A, Shariff A, Kats M, Fang X, Christiano C, et al. Tweeting the meeting: an in-depth analysis of Twitter activity at Kidney Week 2011. PLoS One 2012;7(7):e40253 [FREE Full text] [doi: 10.1371/journal.pone.0040253] [Medline: 22792254]




XSL•FO
RenderX

18. Wen X, Parra D, Trattner C. How groups of people interact with each other on twitter during academic conferences. In: Proceedings of the Companion Publication of the 17th ACM Conference on Computer Supported Cooperative Work & Social Computing. 2014 Presented at: CSCW2014; March 14-18, 2014; Vancouver, BC.
19. Barrat A, Cattuto C, Szomszor M, den Broeck WV, Alani H. Social dynamics in conferences: analyses of data from the live social semantics application. In: The Semantic Web – ISWC 2010. Berlin/Heidelberg: Springer; 2010:17-33.
20. Szomszor M, Kostkova P, Cattuto C, den Broeck WV, Barrat A, Alani H. Providing enhanced social interaction services for industry exhibitors at large medical conferences. In: Developments in E-systems Engineering. 2011 Presented at: DeSE 2011; Dec 6-8, 2011; Dubai.
21. Nature. 2013. SpotOn: tag archives - social media case study URL: http://www.nature.com/spoton/tag/social-media-case-study/ [accessed 2016-02-10] [WebCite Cache ID 6fC8h44dK]
22. TwUnfollow. 2015. URL: http://twunfollow.com [accessed 2016-02-10] [WebCite Cache ID 6fC8kX9RK]
23. de Quincey E, Kostkova P, Jawaheer G. Promoting e-health resources: lessons learned. In: Electronic Healthcare. Berlin/Heidelberg: Springer; 2012:103-110.
24. Isella L, Romano M, Barrat A, Cattuto C, Colizza V, Van den Broeck W, et al. Close encounters in a pediatric ward: measuring face-to-face proximity and mixing patterns with wearable sensors. PLoS One 2011;6(2):e17144 [FREE Full text] [doi: 10.1371/journal.pone.0017144] [Medline: 21386902]
25. Miller L, Johnson C, Weaver A. American Society of Primatologists. 2002. Expanded guidelines for giving a poster presentation URL: https://www.asp.org/education/posterpresentationguidelines.cfm [accessed 2016-05-17] [WebCite Cache ID 6hZcNukvn]
26. Purrington C. colinpurrington.com. Designing conference posters URL: http://colinpurrington.com/tips/academic/posterdesign [accessed 2016-05-17] [WebCite Cache ID 6fC8zrRcz]
27. MacIntosh-Murray A. Poster presentations as a genre in knowledge communication: a case study of forms, norms, and values. Science Communication 2007 Mar 01;28(3):347-376. [doi: 10.1177/1075547006298251]
28. Flickr. URL: https://www.flickr.com/help/galleries/ [accessed 2016-02-10] [WebCite Cache ID 6fC9887uK]
29. CoveritLive. URL: http://www.coveritlive.com/ [accessed 2016-02-10] [WebCite Cache ID 6fC9BKRBy]


## Abbreviations

**API:** application programming interface
**RFID:** radio-frequency identification


*Edited by G Eysenbach; submitted 29.03.15; peer-reviewed by D Paolotti, H Nakada; comments to author 15.08.15; revised version received 10.02.16; accepted 20.02.16; published 19.07.16.*

*Please cite as:*
*Winandy M, Kostkova P, de Quincey E, St Louis C, Szomszor M*
*Follow #eHealth2011: Measuring the Role and Effectiveness of Online and Social Media in Increasing the Outreach of a Scientific Conference*
*J Med Internet Res 2016;18(7):e191*
*URL: http://www.jmir.org/2016/7/e191/*
*doi:10.2196/jmir.4480*
*PMID:27436012*


©Marcel Winandy, Patty Kostkova, Ed de Quincey, Connie St Louis, Martin Szomszor. Originally published in the Journal of Medical Internet Research (http://www.jmir.org), 19.07.2016. This is an open-access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/2.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on http://www.jmir.org/, as well as this copyright and license information must be included.



XSL•FO
RenderX

Original Paper

# How Affiliation Disclosure and Control Over User-Generated Comments Affects Consumer Health Knowledge and Behavior: A Randomized Controlled Experiment of Pharmaceutical Direct-to-Consumer Advertising on Social Media

David Christopher DeAndrea[1], PhD; Megan Ashley Vendemia[1], MA, MBA

School of Communication, The Ohio State University, Columbus, OH, United States

**Corresponding Author:**
David Christopher DeAndrea, PhD
School of Communication
The Ohio State University
154 N Oval Mall
3066 Derby Hall
Columbus, OH, 43210
United States
Phone: 1 614 292 4863
Fax: 1 614 292 2055
Email: deandrea.1@osu.edu

## Abstract

**Background:** More people are seeking health information online than ever before and pharmaceutical companies are increasingly marketing their drugs through social media.

**Objective:** The aim was to examine two major concerns related to online direct-to-consumer pharmaceutical advertising: (1) how disclosing an affiliation with a pharmaceutical company affects how people respond to drug information produced by both health organizations and online commenters, and (2) how knowledge that health organizations control the display of user-generated comments affects consumer health knowledge and behavior.

**Methods:** We conducted a 2×2×2 between-subjects experiment (N=674). All participants viewed an infographic posted to Facebook by a health organization about a prescription allergy drug. Across conditions, the infographic varied in the degree to which the health organization and commenters appeared to be affiliated with a drug manufacturer, and the display of user-generated comments appeared to be controlled.

**Results:** Affiliation disclosure statements on a health organization's Facebook post increased perceptions of an organization-drug manufacturer connection, which reduced trust in the organization (point estimate –0.45, 95% CI –0.69 to –0.24) and other users who posted comments about the drug (point estimate –0.44, 95% CI –0.68 to –0.22). Furthermore, increased perceptions of an organization-manufacturer connection reduced the likelihood that people would recommend the drug to important others (point estimate –0.35, 95% CI –0.59 to –0.15), and share the drug post with others on Facebook (point estimate –0.37, 95% CI –0.64 to –0.16). An affiliation cue next to the commenters' names increased perceptions that the commenters were affiliated with the drug manufacturer, which reduced trust in the comments (point estimate –0.81, 95% CI –1.04 to –0.59), the organization that made the post (point estimate –0.68, 95% CI –0.90 to –0.49), the likelihood of participants recommending the drug (point estimate –0.61, 95% CI –0.82 to –0.43), and sharing the post with others on Facebook (point estimate –0.63, 95% CI –0.87 to –0.43). Cues indicating that a health organization removed user-generated comments from a post increased perceptions that the drug manufacturer influenced the display of the comments, which negatively affected trust in the comments (point estimate –0.35, 95% CI –0.53 to –0.20), the organization (point estimate –0.31, 95% CI –0.47 to –0.17), the likelihood of recommending the drug (point estimate –0.26, 95% CI –0.41 to –0.14), and the likelihood of sharing the post with others on Facebook (point estimate –0.28, 95% CI –0.45 to –0.15). (All estimates are unstandardized indirect effects and 95% bias-corrected bootstrap confidence intervals.)

**Conclusions:** Concern over pharmaceutical companies hiding their affiliations and strategically controlling user-generated comments is well founded; these practices can greatly affect not only how viewers evaluate drug information online, but also how likely they are to propagate the information throughout their online and offline social networks.



XSL•FO
RenderX

*(J Med Internet Res 2016;18(7):e189)* doi:10.2196/jmir.5972

**KEYWORDS**

direct-to-consumer advertising; DTCA; social media; Facebook; pharmaceutical marketing; online promotion of prescription drugs; health communication

## Introduction

The emergence of new interactive communication media has drastically affected the way many people seek out health information and discuss health topics. More than 70% of Internet users seek health information online for themselves and others; 55% of all users go online to diagnose ailments, 40% go online to seek information about medical treatments, and over 15% go online to look up drugs that they saw advertised [1]. Given the amount of people who use the Internet for health information seeking, it is not surprising that pharmaceutical companies have increasingly marketed their drugs through interactive websites and social media [2,3]; this practice is commonly referred to as direct-to-consumer advertising (DTCA). Recent estimates indicate that online pharmaceutical DTCA expenditures have continued to increase, while DTCA spending through more traditional broadcast media (eg, TV) has decreased. Due to its unparalleled scope and a complete lack of rigid law enforcement, online DTCA contributes to the characterization of Internet activity as being akin to the "Wild West" [4]. Overall, there is a critical need for more research that examines how online DTCA affects consumer health knowledge and decision making [5].

This study examines two serious concerns about online DTCA that researchers and health professionals view as essential to address [5-8]. The first major concern is that pharmaceutical companies might market their drugs on social media indirectly through seemingly neutral third-party sources that are, in reality, controlled or influenced by the pharmaceutical companies [5-7,9]. We examine the ramifications of such practices on social media for both individual commenters and for organizations that post drug information. The second major concern relates to the practice of strategically controlling user-generated contributions. For instance, companies might present "...moderated forums/sites that appear interactive but only offer one-sided communication" (p 824 [9]). Likewise, it is "...possible for manufacturers to support third-party bloggers, posters, and Twitter users who make flattering claims and discredit negative claims about their products in online discussions" (p 2088 [7]).

To understand the magnitude of these concerns, we conducted an experiment that examines how (1) disclosing an affiliation with a drug company, and (2) strategically controlling user-generated comments affects the evaluation of drug information provided on social media. Specifically, we examine the degree to which these concerns and practices affect peoples' trust in multiple information sources, their likelihood of recommending a pharmaceutical drug to friends and family, and their likelihood of sharing pharmaceutical drug information with others in their online social network.

### Direct-to-Consumer Advertising

Although proponents argue that DTCA has benefits, such as educating consumers and improving patient-physician interaction, opponents argue that it has many harmful effects, such as misinforming patients and overemphasizing benefits [10]. The United States and New Zealand are the only two developed countries where DTCA is legal [11]. The US Food and Drug Administration (FDA) is responsible for DTCA oversight and has regulated drug marketing that appears through traditional broadcast media, banning misleading statements and creating categories for different types of advertisements. Product claim ads mention a specific drug by name and the ailment it intends to treat. This type of advertisement must follow a "fair balance" rule, meaning that benefits and risks are given equal coverage. Help-seeking ads, which mention an ailment but not the name of a drug, and reminder ads, which mention a drug but not what it treats, are not required to meet the fair balance rule.

The emergence of social media and other interactive platforms has only exacerbated concerns related to DTCA leading prominent scholars and health professionals to wonder "...whether regulatory responses by FDA are responsive and adaptive enough to address the inherent challenges faced by a universe of digital and Internet-based forms of DTCA" (p 271 [12]). Specifically, researchers and health professionals are extremely concerned that companies will market their drugs online in ways that (1) obscure the role companies play in producing drug information, and (2) strategically control user-generated contributions to promote a favorable, one-sided view of a company and its products [5,8,12]. A recent content analysis clearly documents how prominently pharmaceutical companies are using Facebook, YouTube, and Twitter for promotional activities [3]. In order to better understand how these specific concerns about online DTCA might affect the evaluation of drug information appearing on social media, we draw on research that more broadly examines how features of new media affect the evaluation of information.

### Warranting Theory

Researchers have applied warranting theory [13,14] to understand how features of new media affect evaluations of people [15], companies [16,17], products [18], and websites [19]. A central premise of warranting theory is that people trust information more or less depending on its warranting value; the warranting value of information is defined as the degree to which information is controlled or manipulated by the target it describes. The more people perceive information about a target (eg, person, organization, company) to be under the control of the target, the less they trust the information [20].



XSL•FO
RenderX

### Masking the Identity or Affiliations of an Information Source

As in the case of online DTCA, the complexities of new media can make it difficult to know the true identity of an information source or with whom the source is affiliated. Researchers have documented the many ways that online sources try to influence viewers by masking their true identities [21,22]. The prevalence of fake online reviews, commissioned or produced by the target being evaluated [23], provides a prominent example of how source obfuscation occurs online. Warranting theory predicts that information produced by a third party is more influential the less people perceive it to be under the control of the target being evaluated [13]. A recent study supported this prediction by indicating that online reviews were less impactful the more people were uncertain about the true identity of the reviewers [16].

Overall, consumers tend to trust user-generated reviews or word-of-mouth more than traditional advertisements [24,25]. The persuasive value of personal testimonials in health settings is also well understood [6], and is what makes researchers [9] so concerned about online DTCA wherein the affiliations people or companies have with pharmaceutical companies are not disclosed. Consistent with warranting theory, it is expected that people will trust favorable information about a pharmaceutical drug on social media more the less they perceive the drug manufacturer to be affiliated with third-party information sources. This expectation applies to organizations that post drug information to social media as well as people who comment on such posts. Cues that suggest a connection between organizations or commenters and a drug manufacturer should diminish the warranting value of any favorable evaluations they produce, thus (1) making the information less trustworthy, (2) making people less likely to recommend the drug, and (3) making people less likely to share the information with others in their social network.

### Strategically Controlling User-Generated Posts on Social Media

Even when information sources truly are third parties with no connection to the target they are evaluating, features of new media can still permit targets to exert control over information they produce. Notably, targets can selectively delete some user-generated content to promote a favorable view of themselves or their products. When targets can delete user-generated content, they can exert control over third-party content, not by editing or influencing the content of messages, but by curating the composition of third-party messages that appear online. As such, targets can orchestrate the false appearance that all online commenters or reviewers share the same opinion about some issue or product (eg, they all view a drug favorably). How targets strategically control the dissemination of third-party information is increasingly important to examine because of the enhanced trust people afford user-generated content [26-29]. If no cues exist to suggest that a target is controlling the dissemination of user-generated content, viewers are likely to view the user-generated content as having a high degree of warranting value [13]. However, actions or cues that suggest that a target is controlling the dissemination of user-generated content (eg, deleting comments, restricting access to content) can lower the perceived warranting value of the information, and thus its impact on viewers' attitudes and behaviors. A recent experiment supports this prediction in an online review setting showing that positive reviews of a company led to more favorable attitudes toward the company, the more viewers believed that the company could not control or influence what reviews were displayed [16]. We argue that the same principle should apply to control over user-generated drug evaluations on social media. Specifically, cues that indicate that user-generated evaluations of a drug have been removed from a post should increase perceptions that the drug manufacturer is controlling the dissemination of the user-generated evaluations, thus (1) lowering trust in any remaining favorable user-generated evaluations, (2) making people less likely to recommend the drug, and (3) making people less likely to share the information with others in their social network.

## Methods

### Research Design Overview

Participants in this 2 (organization affiliation) × 2 (commenter affiliation) × 2 (comment deletion) between-subjects experiment were randomly assigned to one of eight conditions. Across all conditions, participants viewed an infographic posted on Facebook by a fictitious health organization about a fictitious prescription allergy drug called OpenAir; the post was always accompanied by user-generated comments. After viewing the stimulus materials, participants completed an online questionnaire and were thanked for their participation.

### Ethics

The study was determined category 2 exempt research by the Ohio State University Institutional Review Board in accordance with the U.S. Department of Health & Human Services. It is most appropriately categorized as research that uses common survey procedures. Authors did not register the trial prospectively because they did not regard this as a clinical trial.

### Sample

The sample consisted of 674 participants from an online panel who received financial compensation in exchange for their participation. Participants ranged in age from 18 to 85 years (mean 52.86, SD 15.00) and identified as Caucasian (n=587), African American (n=44), Asian or Asian American (n=20), Native American (n=3), Hispanic (n=10), and other (n=10). More participants identified as female (n=515) than male (n=159). Participants were recruited by Qualtrics; incentives were handled by Qualtrics and disclosed to participants prior to their participation.

### Stimulus Materials

Across all conditions, participants viewed an infographic that the health organization Expert Opinions in Medicine (EOIM) posted to Facebook. The infographic always contained a quote from a medical doctor about a prescription allergy drug called OpenAir made by Darby Pharmaceuticals. The quote indicated that the drug is effective at treating seasonal allergies. The



XSL•FO
RenderX

infographic was always accompanied by three comments from users who indicated that the drug is very effective. The health organization EOIM, the company Darby Pharmaceuticals, and the prescription drug OpenAir were all fictitious and were created to resemble actual entities.

For the organization affiliation factor, changes were made to the infographic to induce variability in participants' awareness of EOIM's affiliation with Darby Pharmaceuticals—the ostensible manufacturer of the drug OpenAir. In the nonaffiliated condition, the following statement appeared at the bottom of the infographic: "Expert Opinions in Medicine—An Independent Research Organization." In the affiliated condition, the statement read "Expert Opinions in Medicine—A Research Organization Funded by Darby Pharmaceuticals." In addition to this difference, a medical doctor who works for EOIM is quoted in the infographic. The nonaffiliated condition indicated the medical doctor is the Executive Director at EOIM, whereas the affiliated condition adds that he is also an OpenAir Senior Research Scientist at Darby Pharmaceuticals.

We manipulated the other two experimental factors in the comment section that accompanied the infographic. Across all conditions, there were three positive comments on the infographic post. For the commenter affiliation induction, "Darby Pharma" appeared next to the commenters' names in the affiliation condition (eg, Mel L, Stockton-Darby Pharma). In the nonaffiliation condition, no indication was provided that the commenters were associated with Darby Pharmaceuticals. For the dissemination control induction, the caption "We reserve the right to hide or delete comments" appeared above the comments in the deletion condition. In addition, the comment section also indicated that some comments had been hidden. In the nonremoval condition, no statement was provided about the organization's deletion policy and no cues indicated that deletion had occurred. Sample stimuli are provided in Figures 1 and 2.

**Figure 1.** Sample stimulus Facebook infographic post: organization affiliation, comment deletion, and commenter affiliation condition.




**Figure 2.** Sample stimulus Facebook infographic post: no organization affiliation, no comment deletion, and no commenter affiliation condition.




## Measures

### *Organization Affiliation*

The extent to which the organization EOIM was perceived to be affiliated with Darby Pharmaceuticals was assessed using three items measured on 7-point scales with endpoints ranging from extremely unlikely to extremely likely. The items were "Expert Opinions in Medicine is affiliated with Darby Pharmaceuticals," "Expert Opinions in Medicine is connected with Darby Pharmaceuticals," and "Expert Opinions in Medicine is funded by Darby Pharmaceuticals." The reliability of all scales was assessed via Cronbach's alpha ($\alpha$=.94).

### *Dissemination Control*

The extent to which Darby Pharmaceuticals was perceived to control what comments appeared on the Facebook post was assessed with items validated in previous work [30]. Four items measured on 7-point scales were used with endpoints ranging from extremely unlikely to extremely likely ($\alpha$=.97). Items included "Darby Pharmaceuticals controlled what comments appeared on the Facebook post" and "Only comments approved by Darby Pharmaceuticals appeared on the Facebook post."

### *Commenter Affiliation*

The extent to which commenters were perceived to be affiliated with Darby Pharmaceuticals was assessed using three items



XSL•FO
RenderX

measured on 7-point scales with endpoints ranging from extremely unlikely to extremely likely (α=.96). Items included "The people posting comments are affiliated with Darby Pharmaceuticals" and "The people posting comments are employed by Darby Pharmaceuticals."

### *Organization Trust*

Perceived trust in the organization EOIM was assessed via four items on 7-point semantic differential scales. The stem stated "The organization Expert Opinions in Medicine is..." and the bipolar adjectives were as follows: not credible/credible, untrustworthy/trustworthy, not believable/believable, biased/unbiased (α=.96).

### *Comment Trust*

Perceived trust in the comments was assessed via five items on 7-point semantic differential scales. The stem stated "The comments people posted are..." and the bipolar adjectives were as follows: untrustworthy/trustworthy, biased/unbiased, not credible/credible, not reliable/reliable, not believable/believable (α=.97).

### *Drug Recommendation*

To assess interpersonal influence, we measured the extent to which people would recommend the drug OpenAir to important others using four items measured on 7-point scales with endpoints ranging from strongly disagree to strongly agree (α=.97). Items included "I would recommend OpenAir to a friend looking for a good allergy medication" and "I would recommend OpenAir to a family member looking for a good allergy medication."

### *Facebook Endorsement*

To assess influence through mass communication, we measured the extent to which people would endorse and share the post about OpenAir with an entire online social network using four items measured on 7-point scales with endpoints ranging from strongly disagree to strongly agree (α=.98). Items included "I would 'share' the post by Expert Opinions in Medicine with my Facebook friends," "I would 'share' the post by Expert Opinions in Medicine with my Facebook friends who have allergy problems," "I would 'like' the post by Expert Opinions in Medicine on Facebook," and "I would 'like' the organization Expert Opinions in Medicine on Facebook."

### *Demographics*

Whether participants suffer from seasonal allergies was assessed with one item. Participants were asked to indicate how much they agree with the following statement: "Seasonal allergies are a problem for me." Participants were also asked to indicate whether they use Facebook. Facebook users then indicated approximately how often they access Facebook. In addition, participants responded to demographic items including gender, race/ethnicity, and age.

### *Attention Checks*

In all conditions, participants were asked one question for each induction to determine the degree to which they attended to the information in their assigned condition. To check the organization affiliation induction, we asked, "According to the Facebook post you viewed, which of the following statements is true?" The answer options were, "Expert Opinions in Medicine is an independent research organization," "Expert Opinions in Medicine is an organization that is funded by Darby Pharmaceuticals," and "I am not sure." To check the commenter affiliation induction, we asked, "Was there any evidence that the people who commented on the Facebook post were associated with Darby Pharmaceuticals?" and to check the comment deletion induction we asked, "Was there any evidence on the Facebook post you viewed that the organization hides or deletes comments?" The answer options for both questions were yes, no, and I am not sure.

## Results

### Data Exclusion

Before conducting the primary analyses, we examined how participants responded to the attention check questions. Participants who answered incorrectly to one or more of the attention check items were removed from the analyses (n=265). Participants who did not answer any of the questions incorrectly were retained (n=409). These two groups of participants did not significantly differ in their age, gender, ethnicity, Facebook use, or whether they suffered from seasonal allergies (all *P* values >.21).

### Analysis Plan

For each experimental factor, we first provide a *t* test that directly estimated how each induction affected perceptions of the mediating construct it was expected to vary (ie, organization affiliation, dissemination control, or commenter affiliation). Next, the macro PROCESS [31] was used to estimate the indirect effect each induction had on the outcome measures, through the proposed mediator. We first provide unadjusted indirect effect estimates and then provide covariate-adjusted estimates. A zero-order correlation matrix is provided that includes means and standard deviations for all variables in the analyses (Table 1).



XSL•FO
RenderX

**Table 1.** Zero-order correlations, means, and standard deviations (N=409).

| Variable | Pearson *r* | | | | | | | Mean (SD) |
|---|---|---|---|---|---|---|---|---|
| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | |
| 1. Organization affiliation | 1 | .41** | .39** | –.31** | –.26** | –.21** | –.16* | 5.44 (1.59) |
| 2. Dissemination control | | 1 | .57** | –.48** | –.53** | –.41** | –.40** | 5.40 (1.74) |
| 3. Commenter affiliation | | | 1 | –.42** | –.50** | –.38** | –.33** | 4.74 (1.76) |
| 4. Organization trust | | | | 1 | .80** | .69** | .66** | 3.70 (1.58) |
| 5. Comment trust | | | | | 1 | .68** | .65** | 3.55 (1.58) |
| 6. Drug Recommendation likelihood | | | | | | 1 | .78** | 4.07 (2.15) |
| 7. Facebook endorsement | | | | | | | 1 | 2.86 (1.60) |

* *P*<.01, ** *P*<.001

## Organization Affiliation

An independent samples *t* test indicated that participants perceived the organization EOIM as more affiliated with Darby Pharmaceuticals when the infographic provided a funding disclosure relative to when it did not ($t_{407}$=12.23, *P*<.001, $\eta^2$=.27). Using Model 4 of the macro PROCESS, we estimated the indirect effect the induction had through perceptions of organizational affiliation on each outcome measure; each estimate is provided with its corresponding 95% bias-corrected bootstrap confidence interval based on 10,000 resamples. Significant indirect effects were found on organization trust (point estimate –0.45, 95% CI –0.69 to –0.24), comment trust (point estimate –0.44, 95% CI –0.68 to –0.22), drug recommendation likelihood (point estimate –0.35, 95% CI –0.59 to –0.15), and Facebook endorsement likelihood (point estimate –0.37, 95% CI –0.64 to –0.16). The preceding analyses were reran controlling for perceptions of dissemination control (ie, Darby Pharmaceuticals controlled what comments appeared) and perceptions that the commenters were affiliated with Darby Pharmaceuticals. When including the covariates in the mediation models, all of the indirect effects became nonsignificant.

To further probe the nature of the relationships, we conducted tests of moderated mediation using Model 14 in PROCESS. The significant indirect effects reported previously were not moderated by participants' perceptions that the commenters were affiliated with Darby Pharmaceuticals. However, perceptions of dissemination control did moderate three of the significant indirect effects. The overall pattern indicated that—when perceptions of dissemination control were low—the organizational affiliation induction did not indirectly affect (1) trust in the organization, (2) trust in the comments, and (3) the likelihood of recommending the drug. However, as perceptions of dissemination control increased, the size of the indirect effects significantly increased to a substantive degree. For each outcome measure, indirect effects were estimated at three levels of the moderating variable (ie, dissemination control): 1 SD below the mean, at the mean, and 1 SD above the mean. The point estimates were as follows for each outcome: organization trust (point estimates –0.09, –0.29, –0.47), comment trust (point estimates –0.02, –0.12, –0.26), and drug recommendation likelihood (point estimates 0.01, –0.14, –0.29). The direct test of moderated mediation for each outcome was as follows: organization trust index (point estimate –0.11, 95% CI –0.21 to –0.02), comment trust index (point estimate –0.09, 95% CI –0.18 to –0.002), and drug recommendation likelihood index (point estimate –0.09, 95% CI –0.18 to –0.004). A similar pattern was found for the Facebook endorsement outcome but the 95% confidence interval narrowly included zero: index (point estimate –0.08, 95% CI –0.19 to 0.01).

## Dissemination Control

An independent samples *t* test indicated that participants perceived Darby Pharmaceuticals to have more control over what comments appeared on the EOIM post when cues indicated that deletion had occurred ($t_{407}$=4.39, *P*<.001, $\eta^2$=.05). Using Model 4 of the macro PROCESS, we estimated the indirect effect the induction had through perceptions of dissemination control on each outcome measure. Significant indirect effects were found on organization trust (point estimate –0.31, 95% CI –0.47 to –0.17), comment trust (point estimate –0.35, 95% CI –0.53 to –0.20), drug recommendation likelihood (point estimate –0.26, 95% CI –0.41 to –0.14), and Facebook endorsement likelihood (point estimate –0.28, 95% CI –0.45 to –0.15). The preceding analyses were reran controlling for perceptions that the organization and the commenters were affiliated with Darby Pharmaceuticals. When including the covariates, all the indirect effects remained significant; the confidence interval of each adjusted estimate overlapped with its respective nonadjusted confidence interval. That is, the estimates did not significantly differ in magnitude.

## Commenter Affiliation

An independent samples *t* test indicated that participants perceived the commenters to be more affiliated with Darby Pharmaceuticals when the cue "Darby Pharma" appeared next to their names ($t_{407}$=10.59, *P*<.001, $\eta^2$=.22). Using Model 4 of the macro PROCESS, we estimated the indirect effect the induction had through perceptions of commenter affiliation on each outcome measure. Significant indirect effects were found on organization trust (point estimate –0.68, 95% CI –0.90 to –0.49), comment trust (point estimate –0.81, 95% CI –1.04 to –0.59), drug recommendation likelihood (point estimate –0.61, 95% CI –0.82 to –0.43), and Facebook endorsement likelihood (point estimate –0.63, 95% CI –0.87 to –0.43). The preceding analyses were reran controlling for perceptions that the organization was affiliated with Darby Pharmaceuticals and



XSL•FO
RenderX

perceptions that Darby Pharmaceuticals controlled the dissemination of the comments. When including the covariates, all the indirect effects remained significant, the confidence interval of each adjusted estimate always overlapped with its respective nonadjusted confidence interval except in one case. The covariate-adjusted indirect effect on comment trust was significantly attenuated (point estimate –0.37, 95% CI –0.54 to –0.23).

### Full Sample

As noted previously, participants were removed from the analyses if they incorrectly responded to one of the attention check items. The purpose of removing the participants was to reduce error and provide a clearer test of the hypothesized relationships. However, in real-world settings, people may only provide fleeting attention to social media posts and/or may be unable to accurately recall what they viewed. As such, there is some value in being exhaustive and looking at the estimates for the full sample—even if this includes participants who made no honest attempt to read or respond to the survey items. As indicated in Table 2, all the indirect effects were significant for the full sample. Although the estimates for the full sample are attenuated relative to the reduced sample, they cannot be statistically differentiated because their respective 95% confidence intervals overlap.

**Table 2.** Indirect effects full sample (N=672): point estimates and 95% confidence intervals.[a]

| Factor | Dependent variables, point estimate (95% CI) | | | |
|---|---|---|---|---|
| | Organization trust | Comment trust | Drug recommendation likelihood | Facebook endorsement |
| Organization affiliation | –0.32 (–0.46, –0.19) | –0.26 (–0.40, –0.13) | –0.18 (–0.31, –0.06) | –0.16 (–0.30, –0.03) |
| Comment deletion | –0.18 (–0.30, –0.06) | –0.21 (–0.34, –0.07) | –0.16 (–0.27, –0.06) | –0.17 (–0.29, –0.06) |
| Commenter affiliation | –0.47 (–0.62, –0.34) | –0.56 (–0.73, –0.42) | –0.42 (–0.56, –0.30) | –0.40 (–0.55, –0.28) |

[a] Estimates are provided with their respective 95% bias-corrected bootstrap confidence interval based on 10,000 resamples.

## Discussion

### Principal Findings

The findings from this study illustrate how important it is to better understand the effects of DTCA in a new media environment. A major concern expressed in past research [5,7,9] is that information sources might be “blurred” online, making it difficult to know when a pharmaceutical company is sponsoring or influencing the production of drug information. The results of this study suggest that cues that disclose connections between health organizations and pharmaceutical companies affect how people process drug information posted on social media. Specifically, disclosing an affiliation decreased (1) trust in an organization that posted information about a drug, (2) trust in comments posted by other site users about the drug, (3) the likelihood of recommending the drug to family or friends, and (4) the likelihood of propagating the drug message further throughout their online social network. Illustrating the complexity of new media environments that contain multiple information sources, the results also indicate that these effects are increasingly pronounced when it appears that a website proprietor controls the dissemination of user-generated comments on a webpage.

Beyond moderating the effect perceptions of organizational affiliation had on the outcome measures, perceptions of control over the dissemination of user-generated content independently affected the outcomes. This type of strategic control over user-generated content has been emphasized as a major concern for online DTCA [7,9]. The findings validate these concerns and help estimate how greatly controlling the dissemination of user-generated content can affect people who view health information posted online. Cues that indicated that an organization removed some of the user-generated comments that accompanied their posts increased people’s perceptions that the drug manufacturer was behind the removal. Notably, the more people thought that the drug manufacturer was controlling the dissemination of the user-generated comments, the less people trusted the user comments and the health organization that posted the infographic. Again, the complexities of a new media environment are illustrated. How information posted by website users is perceived to be controlled not only affects how people evaluate remaining user contributions, it also affects how people view the proprietor of the website (eg, the health organization EOIM). In addition to influencing how people trusted the user-generated comments and the health organization EOIM, control over the dissemination of user-generated content affected the likelihood that people would recommend the drug to others, and endorse/share the information with others in their online social network.

The commenter affiliation induction had similar, yet independent, effects on all the outcome measures. The results indicate that a single affiliation cue next to commenters’ names can significantly increase people’s knowledge that the commenters are affiliated with the drug company, which, in turn, can affect trust in a health organization, comment trust, drug recommendation likelihood, and the likelihood of endorsing and sharing the information with others online. Because personal testimonials from average citizens are highly influential, paid representatives or company employees who post information online without disclosing their connection to a pharmaceutical company is thought to be exceedingly troublesome [9]. The findings from this study distinctly illustrate how impactful it can be when company-affiliated individuals masquerade as neutral, third-party contributors online.

From a practical perspective, the results highlight the need for future FDA guidelines to mandate that pharmaceutical companies clearly disclose connections within messages posted



XSL•FO
RenderX

to any website or social media platform that they directly fund, control, or support in some manner. Past research suggests companies have minimized or obscured such disclosures [32]. The bigger challenge, however, is whether it is feasible to regulate how pharmaceutical companies control information about their products across new media platforms that they might only indirectly influence or control [7,12]. Obviously pharmaceutical companies should not be expected to police the entire Internet. Nevertheless, regulations that only pertain to content posted to "official" company media might overlook relationships that—according to the results of this study—would greatly influence how people evaluate drug information.

The FDA has issued draft guidance documents that outline (1) when companies are responsible for user-generated content, (2) recommendations for how they should respond to user-generated content for which they are not required to respond, and (3) recommendations for how to convey risk information when online platforms have space or characters limitations [33,34]. Notably, the documents do not directly discuss how companies should respond to online platforms that make it difficult for content contributors to clearly disclose company affiliations and the precise nature of any affiliations. Ironically, the guidance document on user-generated contributions only provides guidance for handling user-generated content that comes from independent, nonaffiliated sources and exists on platforms in which companies have not edited or removed any third-party content. "A firm is thus responsible for communications on the Internet and Internet-based platforms, such as social media, made by its employees or any agents acting on behalf of the firm to promote the firm's product, and these communications must comply with any applicable regulatory requirements" (pp 3-4 [33]). It is helpful that the guidance document confirms that existing regulations apply to content produced by company-affiliated sources and content that exists on platforms over which companies exert control. However, the problem remains that features of many prominent social media sites, such as Facebook, make it easy to mask the identity of an information source and difficult to ascertain whether companies are removing user-generated contributions.

## Theoretical Implications

The findings of this study help extend the explanatory and predictive power of warranting theory by demonstrating how the core theoretical propositions accurately explain and predict phenomena in a new context. Unlike previous tests of warranting theory that have directly examined uncertainty about the true identity of an information source online [16], this study more directly varied and measured the degree to which information sources were affiliated with the target being evaluated. Although the distinction may appear to be minor, uncertainty about the true identity of a third-party source might affect perceptions of warranting value differently relative to perceptions that a third-party source is affiliated with the target of an evaluation. For instance, an unknown source could actually be the target, someone affiliated with the target, or someone unaffiliated with the target. Future research might seek to further explore how these two considerations about the identity of a third-party source relate to one another and affect evaluations of information appearing online.

A notable finding from this work that has novel theoretical implications is that perceptions of organizational affiliation affected the outcomes differently than perceptions of commenter affiliation. It is possible that the differential effects might be attributable to how the constructs were operationalized in this study. However, it is also possible that perceived affiliations between individuals and targets might function differently than perceived affiliations between organizations/companies and targets. If the ceiling for trusting an online commenter is greater than the ceiling for trusting an organization, differential effects might be expected. Future research might seek to explore these possibilities more directly.

## Limitations

Limitations common to experimental research apply to this study. Although the results support theoretically predicted relationships, future research might seek to further the generalizability of the findings. For instance, researchers can seek to examine the effects of online DTCA across different populations, with different drug messages, and on different social media platforms. Researchers might also seek to examine how consumers' general skepticism toward pharmaceutical marketing can moderate the effects found in this study.

A limitation specific to this study is that the attention check questions may have been overly sensitive. For instance, some participants viewed an infographic that indicated that EOIM was "An Independent Research Organization." Participants who indicated that EOIM was not an independent organization and instead was funded by Darby Pharmaceuticals were removed. However, in some of the EOIM nonaffiliation conditions, comments were removed and commenters were affiliated with Darby Pharmaceuticals. It is possible that participants interpreted this combination of cues as indicating EOIM was not really independent, despite the claim that was made. We provide results for the full sample to overcome this limitation, but future researchers should consider how multiple cues might operate in conjunction when seeking to include attention check items designed to reduce measurement error.

## Conclusions

Pharmaceutical companies will seek to market their drugs through whatever media people regularly consume. In the current media landscape, this means drug marketing will occur through social media and online platforms that are interactive and include information from multiple sources. Any attempt to regulate online DTCA needs to thoroughly consider the unique affordances and characteristics of emerging communication technology. Whether regulations can keep pace with advances in communication technology remains to be seen. However, the results of this study provide clear evidence that obscuring (1) the true identity of an information source, (2) the affiliations of an information source, and (3) control over user-generated content can greatly influence consumer health knowledge and behavior.



XSL•FO
RenderX

## Acknowledgments

The authors would like to thank Edward A. Weeks and Molly Mao for their helpful feedback. The lead author would like to thank Noemi for her wisdom and support.

## Conflicts of Interest

None declared.

## References

1. Fox S, Duggan M. Health Online 2013. Washington, DC: Pew Internet & American Life Project; 2013 Jan 15. URL: http://www.pewinternet.org/files/old-media/Files/Reports/PIP_HealthOnline.pdf [accessed 2016-06-22] [WebCite Cache ID 6WKEp8JDd]
2. Mackey TK, Cuomo RE, Liang BA. The rise of digital direct-to-consumer advertising?: Comparison of direct-to-consumer advertising expenditure trends from publicly available data sources and global policy implications. BMC Health Serv Res 2015 Jun;15:236 [FREE Full text] [doi: 10.1186/s12913-015-0885-1] [Medline: 26084705]
3. Tyrawski J, DeAndrea DC. Pharmaceutical companies and their drugs on social media: a content analysis of drug information on popular social media sites. J Med Internet Res 2015 Jun;17(6):e130 [FREE Full text] [doi: 10.2196/jmir.4357] [Medline: 26032738]
4. Perlroth N, Sanger D. The New York Times. 2015 Feb 13. Obama calls for new cooperation to wrangle the 'wild west' internet URL: http://www.nytimes.com/2015/02/14/business/obama-urges-tech-companies-to-cooperate-on-internet-security.html [accessed 2016-06-22] [WebCite Cache ID 6hSBGd0NB]
5. Southwell BG, Rupert DJ. Future challenges and opportunities in online prescription drug promotion research. Comment on: “Trouble spots in online direct-to-consumer prescription drug promotion: a content analysis of FDA warning letters”. Int J Health Policy Manag 2016 Jan;5(3):211-213 [FREE Full text] [doi: 10.15171/ijhpm.2016.05] [Medline: 26927597]
6. Gibson S. Regulating direct-to-consumer advertising of prescription drugs in the digital age. Laws 2014 Jul 09;3(3):410-438. [doi: 10.3390/laws3030410]
7. Greene JA, Kesselheim AS. Pharmaceutical marketing and the new social media. N Engl J Med 2010 Nov 25;363(22):2087-2089. [doi: 10.1056/NEJMp1004986] [Medline: 21105789]
8. Kim H. Trouble spots in online direct-to-consumer prescription drug promotion: a content analysis of FDA warning letters. Int J Health Policy Manag 2015 Aug;4(12):813-821 [FREE Full text] [doi: 10.15171/ijhpm.2015.157] [Medline: 26673465]
9. Liang BA, Mackey T. Direct-to-consumer advertising with interactive internet media: global regulation and public health issues. JAMA 2011 Feb 23;305(8):824-825. [doi: 10.1001/jama.2011.203] [Medline: 21343583]
10. Ventola CL. Direct-to-consumer pharmaceutical advertising: therapeutic or toxic? Pharm Ther 2011 Oct;36(10):669-684 [FREE Full text] [Medline: 22346300]
11. Greene JA, Watkins ES. The vernacular of risk--rethinking direct-to-consumer advertising of pharmaceuticals. N Engl J Med 2015 Sep 17;373(12):1087-1089. [doi: 10.1056/NEJMp1507924] [Medline: 26287749]
12. Mackey TK. Digital direct-to-consumer advertising: a perfect storm of rapid evolution and stagnant regulation. Comment on “Trouble spots in online direct-to-consumer prescription drug promotion: a content analysis of FDA warning letters”. Int J Health Policy Manag 2016 Feb 03;5(4):271-274. [doi: 10.15171/ijhpm.2016.11]
13. DeAndrea DC. Advancing warranting theory. Commun Theor 2014 Apr 08;24(2):186-204. [doi: 10.1111/comt.12033]
14. Walther JB, Parks MR. Cues filtered out, cues filtered in: computer-mediated communication relationships. In: Knapp M, Daly J, editors. Handb Interpers Commun. Thousand Oaks, CA: Sage; 2002:443-479.
15. Fox J, Warber KM, Makstaller DC. The role of Facebook in romantic relationship development: An exploration of Knapp's relational stage model. J Soc Pers Relat 2013 Jan 06;30(6):771-794. [doi: 10.1177/0265407512468370]
16. DeAndrea DC, Van Der Heide B, Vendemia MA, Vang MH. How people evaluate online reviews. Commun Res 2015 Feb 25:1-18. [doi: 10.1177/0093650215573862]
17. Lillqvist E, Louhiala-Salminen L. Facing Facebook: impression management strategies in company-consumer interactions. J Bus Tech Commun 2013 Sep 26;28(1):3-30. [doi: 10.1177/1050651913502359]
18. Johnson BK, Vang MH, Van Der Heide B. Show me the goods. J Media Psychol 2015 Jan;27(1):3-10. [doi: 10.1027/1864-1105/a000126]
19. DeAndrea DC, Van Der Heide B, Easley N. How modifying third-party information affects interpersonal impressions and the evaluation of collaborative online media. J Commun 2014 Dec 17;65(1):62-78. [doi: 10.1111/jcom.12139]
20. Walther JB, Van Der Heide B, Hamel LM, Shulman HC. Self-generated versus other-generated statements and impressions in computer-mediated communication: a test of warranting theory using Facebook. Commun Res 2009 Jan 22;36(2):229-253. [doi: 10.1177/0093650208330251]
21. Hancock J, Guillory J. Deception with technology. In: Sundar S, editor. The handbook of the psychology of communication technology. Malden, MA: Wiley-Blackwell; 2015:270-289.
22. Malbon J. Taking fake online consumer reviews seriously. J Consum Policy 2013 Jan 23;36(2):139-157. [doi: 10.1007/s10603-012-9216-7]



XSL•FO
RenderX

23. Luka M, Zervas G. Harvard Business School Negotiation, Organizations and Markets Unit, Research Paper Series. Boston, MA: Harvard Business School; 2015 Jul 20. Fake it till you make it: reputation, competition, and Yelp review fraud URL: http://people.hbs.edu/mluca/fakeittillyoumakeit.pdf [accessed 2016-06-22] [WebCite Cache ID 6iSMLUt3r]
24. Bughin J, Doogan J, Vetvick O. McKinsey Quarterly. 2010. A new way to measure word-of-mouth marketing URL: http://www.mckinsey.com/business-functions/marketing-and-sales/our-insights/a-new-way-to-measure-word-of-mouth-marketing [accessed 2016-06-22] [WebCite Cache ID 6iSMTbhWd]
25. Nielsen. 2015 Sep 28. Global trust in advertising URL: http://www.nielsen.com/us/en/insights/reports/2015/global-trust-in-advertising-2015.html [accessed 2016-06-22] [WebCite Cache ID 6hSBPY9Bq]
26. Brown J, Broderick AJ, Lee N. Word of mouth communication within online communities: Conceptualizing the online social network. J Interact Mark 2007 Jan;21(3):2-20. [doi: 10.1002/dir.20082]
27. Lee E, Shin SY. When do consumers buy online product reviews? Effects of review quality, product type, and reviewer's photo. Comput Hum Behav 2014 Feb;31:356-366. [doi: 10.1016/j.chb.2013.10.050]
28. Senecal S, Nantel J. The influence of online product recommendations on consumers' online choices. J Retailing 2004 Jan;80(2):159-169. [doi: 10.1016/j.jretai.2004.04.001]
29. Walther JB, Jang J. Communication processes in participatory websites. J Comput-Mediat Comm 2012 Oct 10;18(1):2-15. [doi: 10.1111/j.1083-6101.2012.01592.x]
30. DeAndrea DC, Carpenter CJ. Measuring the construct of warranting value and testing warranting theory. Commun Res 2016 Apr 25:1-23. [doi: 10.1177/0093650216644022]
31. Hayes A. Introduction to Mediation, Moderation, and Conditional Process Analysis: A Regression-Based Approach. New York: Guilford Press; 2013.
32. Sheehan K. Direct-to-consumer (DTC) branded drug web sites risk presentation and implications for public policy. J Advertising 2007;36(3):123-135 [FREE Full text] [doi: 10.2753/JOA0091-3367360310]
33. US Department of Health and Human Services, Food and Drug Administration, Center for Drug Evaluation and Research, Center for Biologics Evaluation and Research, Center for Veterinary Medicine, Center for Devices and Radiological Health. Guidance for Industry: Internet/Social Media Platforms: Correcting Independent Third-Party Misinformation About Prescription Drugs and Medical Devices. 2014 Jun. URL: http://www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/UCM401079.pdf [accessed 2016-06-22] [WebCite Cache ID 6X8QCChgD]
34. US Department of Health and Human Services, Food and Drug Administration, Center for Drug Evaluation and Research, Center for Biologics Evaluation and Research, Center for Veterinary Medicine, Center for Devices and Radiological Health. Guidance for Industry Internet/Social Media Platforms with Character Space Limitations—Presenting Risk and Benefit Information for Prescription Drugs and Medical Devices. 2014 Jun. URL: http://www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/UCM401087.pdf [accessed 2016-06-22] [WebCite Cache ID 6X8QFAUAl]


## Abbreviations

**DTCA:** direct-to-consumer advertising
**EOIM:** Expert Opinions in Medicine
**FDA:** Food and Drug Administration


*Edited by G Eysenbach; submitted 12.05.16; peer-reviewed by C Adams, BL Rollins; comments to author 01.06.16; revised version received 13.06.16; accepted 13.06.16; published 19.07.16.*

*Please cite as:*
*DeAndrea DC, Vendemia MA*
*How Affiliation Disclosure and Control Over User-Generated Comments Affects Consumer Health Knowledge and Behavior: A Randomized Controlled Experiment of Pharmaceutical Direct-to-Consumer Advertising on Social Media*
*J Med Internet Res 2016;18(7):e189*
*URL: http://www.jmir.org/2016/7/e189/*
*doi:10.2196/jmir.5972*
*PMID:27435883*


©David Christopher DeAndrea, Megan Ashley Vendemia. Originally published in the Journal of Medical Internet Research (http://www.jmir.org), 19.07.2016. This is an open-access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/2.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete



XSL•FO
RenderX

bibliographic information, a link to the original publication on http://www.jmir.org/, as well as this copyright and license information must be included.



XSL•FO
RenderX

Original Paper

# Estimating Influenza Outbreaks Using Both Search Engine Query Data and Social Media Data in South Korea

Hyekyung Woo[1], MPH, PhD; Youngtae Cho[1], PhD; Eunyoung Shim[1], MPH, PhD; Jong-Koo Lee[2], MD, PhD; Chang-Gun Lee[3], PhD; Seong Hwan Kim[4], PhD

[1]Department of Health Science and Service, School of Public Health, Seoul National University, Seoul, Republic Of Korea

[2]Global Medical Center, School of Medicine, Seoul National University, Seoul, Republic Of Korea

[3]Real-time Ubiquitous System Laboratory, Department of Computer Science and Engineering, Seoul National University, Seoul, Republic Of Korea

[4]Search SU, Datamining Team, Daum Kakao Incorporated, Seoul, Republic Of Korea

**Corresponding Author:**
Youngtae Cho, PhD
Department of Health Science and Service
School of Public Health
Seoul National University
1 Kwanakro
Kwanakgu
Seoul, 151-172
Republic Of Korea
Phone: 82 10 7135 4610
Fax: 82 2 762 9105
Email: youngtae@snu.ac.kr

## Abstract

**Background:** As suggested as early as in 2006, logs of queries submitted to search engines seeking information could be a source for detection of emerging influenza epidemics if changes in the volume of search queries are monitored (infodemiology). However, selecting queries that are most likely to be associated with influenza epidemics is a particular challenge when it comes to generating better predictions.

**Objective:** In this study, we describe a methodological extension for detecting influenza outbreaks using search query data; we provide a new approach for query selection through the exploration of contextual information gleaned from social media data. Additionally, we evaluate whether it is possible to use these queries for monitoring and predicting influenza epidemics in South Korea.

**Methods:** Our study was based on freely available weekly influenza incidence data and query data originating from the search engine on the Korean website Daum between April 3, 2011 and April 5, 2014. To select queries related to influenza epidemics, several approaches were applied: (1) exploring influenza-related words in social media data, (2) identifying the chief concerns related to influenza, and (3) using Web query recommendations. Optimal feature selection by least absolute shrinkage and selection operator (Lasso) and support vector machine for regression (SVR) were used to construct a model predicting influenza epidemics.

**Results:** In total, 146 queries related to influenza were generated through our initial query selection approach. A considerable proportion of optimal features for final models were derived from queries with reference to the social media data. The SVR model performed well: the prediction values were highly correlated with the recent observed influenza-like illness ($r$=.956; $P$<.001) and virological incidence rate ($r$=.963; $P$<.001).

**Conclusions:** These results demonstrate the feasibility of using search queries to enhance influenza surveillance in South Korea. In addition, an approach for query selection using social media data seems ideal for supporting influenza surveillance based on search query data.

***(J Med Internet Res 2016;18(7):e177)*** doi:10.2196/jmir.4955

**KEYWORDS**

influenza; surveillance; population surveillance; infodemiology; infoveillance; Internet search; query; social media; big data; forecasting; epidemiology; early response



XSL•FO
RenderX

## Introduction

An early and now well-known example of utilizing Internet data for a health-related applications came from the estimation of influenza incidence using anonymous logs of Web search engine queries. First proposed in 2006 by Eysenbach under the umbrella term "infodemiology", numerous recent studies have added further evidence of a correlation between search query data from Google [1-3], Yahoo! [4], Baidu [5], or other medical websites [6] and traditional data used for influenza surveillance, such as influenza-like illness (ILI) and/or laboratory-confirmed data. These studies indicate that individuals faced with disease or ill health will search for information on the Internet regarding their state of health and possible countermeasures to illness; logs of queries submitted to search engines by individuals seeking this information are potential sources of information for detecting emerging epidemics, as it is possible to track changes in the volumes of specific search queries. However, the recent errors arising from Google Flu Trends, which has been predominantly used in previous studies, serves as a reminder to investigators that this novel data paradigm calls for critical assessment and the development of more empirical methodologies to explore the predictive utility of big data [7,8]. It is clear that current and future studies need to focus on methods to more precisely identify the particular phases associated with influenza epidemics based on data from these highly informative sources.

Selecting the queries that are most likely to be associated with influenza epidemics poses a particular challenge for the generation of improved predictions. In previous studies, researchers have utilized queries selected by various methods, such as specific keyword tools offered by particular websites [5], surveys of patients who visited the emergency room [1,9], or common knowledge about influenza including the definition of ILI [9,10], as well as fully automated methods for identifying queries related to influenza from search logs [3,4,6]. Because researchers do not have full access to search logs, an approach using social media data may also be helpful for obtaining information for query selection. Recently, social media data have been highlighted as an additional potential data source for disease surveillance because they contain a greater variety of contextual health information with diverse descriptions of health states. Thus, it could be a useful reference point for researchers who wish to select initial target queries in query-based prediction.

In South Korea, there is currently no forecasting system for infectious disease based on search query data [1,9], despite the high availability and use of the Internet in Korea [11]. Moreover, few studies thus far have evaluated whether such data could be of value in national influenza forecasting [1,9], and a recent study has suggested that Google Trends in the Korean language is insufficient for use as a model for influenza prediction in South Korea [1]. We need to proactively determine whether queries of search engines that are more widely used by Koreans have the capacity to enhance traditional influenza surveillance systems in South Korea. We consider the use of social media data to select queries that are most likely to be associated with influenza epidemics in a situation involving limited access to search logs. An attempt to exploit the complementary nature of two types of data sources could result in a rapid and efficient prediction of the occurrence of influenza and their proliferation, thereby allowing for better recognition of influenza and initiation of preventive measures.

The purpose of this study was to further explore two concerns: (1) to describe a methodological extension for detecting influenza outbreaks using search query data, providing a new approach for query selection through the exploration of contextual information obtained from social media data, and (2) evaluate whether it is possible to use these queries for monitoring influenza epidemics in South Korea.

## Methods

### Data Sources

#### *Epidemiological Surveillance Data*

National influenza surveillance data were obtained from the Korean Center for Disease Control and Prevention (KCDC), which routinely collects epidemiological data and national statistics pertaining to influenza incidence, typically with a 1-week reporting lag [12]. We used clinical data and virological data from April 3, 2011 (listed as week 32) to April 5, 2014 (listed as week 14). For clinical data, we used the rates of physician visits for ILI; for virological data, the rates for positive results for the influenza virus in laboratory tests. The data obtained were anonymous and publicly available.

#### *Social Media Data*

In developing an approach for query selection, we drew on social media data. Social media data were collected from the daily Naver blog (a weblog service offered by the biggest portal site in South Korea [13]) and Twitter posts between September 1, 2010 and August 31, 2013 (3 years), using the social "big data" mining system, SOCIALmetricsAcademy. This system contains social media data crawlers that collect posts from Twitter and the Naver blog. The system also processes text using state-of-the-art natural language processing and text-mining technologies. The Twitter crawler utilizes a streaming application program interface (API) for data collection using the "track keywords" function. We tracked several thousand keywords that were empirically selected and tuned to maximize the coverage of the crawler operating in near real time. We estimated that the daily coverage of the Twitter crawler was more than 80%. The collected posts were fed into a spam-filtering module that checked for posts containing spam keywords written by known spammers. The lists of spam keywords and spammers were semiautomatically monitored and managed. The Naver blog crawler resembles general-purpose Web crawlers, the main difference being that a list of active bloggers for post collection is maintained and automatically expanded. The estimated coverage of the Naver blog crawler was also more than 80%. We applied an extensive spam-filtering process similar to that of the Twitter crawler on the collected blog posts.

The authors and data mining company conducted the search according to the Twitter and blogging website terms and



XSL•FO
RenderX

conditions of use. All Twitter and Naver blog posts were publicly available and the information collected did not reveal the identity of the social media users; thus, user confidentiality was preserved.

### *Search Engine Query Data*

The query data originated from the search engine on the Korean website, Daum [14]. Although Google is the most-used search engine in the world, it is not dominant in South Korea. Local search engines based on the Korean language, such as Daum, are more widely used than Google. Daum is the second-largest search engine in the portal sites market of South Korea [15]. Because the query data of Korean websites were not publicly available, we sent the list of target queries to Daum and received scaled volume data pertaining to the queries listed. Weekly relative volumes of queries submitted to the search engine between April 3, 2011 and April 5, 2014, were used for analysis. The relative volumes were calculated by dividing the number of each query by the total number of search queries in any given week. The website Daum is written in Korean, thus the submitted queries are primarily in Korean. No information was available that could have potentially revealed the identity of a website visitor; therefore, complete confidentiality was maintained.

## Query Selection

To obtain queries related to influenza that were submitted to the Daum search engine by the Korean population at large, several approaches were applied. Search queries were obtained using the following methods.

### *Seed Keyword for Exploring the Queries*

Although "influenza" is the official term used by the KCDC, *dokgam*, *inpeulruenja*, *peulru*, and *sinjongpeulru* are the words typically used in Korea to describe influenza. Since the 2009 pandemic of influenza virus A (H1N1), the term *sinjongpeulru* to describe the new strain of flu has been more popular in Korea than the term influenza A (H1N1). Thus, *dokgam*, *inpeulruenja*, *peulru*, *sinjongpeulru*, "influenza," and "flu" were defined as seed keywords for exploring the queries. Because Web search queries typically consist of word combinations of an average of two or three terms [16,17], these seed keywords were also used as essential keywords in word combinations.

### *Exploring Influenza-Related Words Through Social Media Data*

To obtain search queries related to influenza, we considered the words that usually appear with the word influenza in the accumulated posts submitted to Twitter and blogs. We first conducted synonym processing for the seed keywords of *dokgam*, *inpeulruenja*, *peulru*, *sinjongpeulru*, influenza, and flu, and named the resulting app Flu. Then, we investigated the words most likely to be associated with Flu using the accumulated posts during the critical 3-year period (between September 1, 2010 and August 31, 2013). Association analysis was performed to identify tuples of topic keyword and associated keywords. This analysis resulted in a total of 157 associated words.

Certain words associated with influenza were not related to influenza seasons or were not commonly entered into search engines. We excluded keywords that occurred infrequently during the influenza season and those that showed nonsequential patterns in the time series throughout the tracking period. Although relatively rare, we also excluded Korean word combinations written in the form of an incomplete sentence. Therefore, we excluded words considered as inadequate candidates for search query following the keyword filtering; in our first phase, we generated 103 candidate queries of single words or word combinations consisting of seed keywords and/or words associated with influenza as determined using social media data.

### *Identifying Chief Concerns Related to Influenza*

Some additional queries related to influenza were obtained through a review of influenza symptoms referring to patients' chief concerns. The influenza surveillance system of the KCDC defines ILI as the sudden onset of high fever (38°C or greater) accompanied by a cough and/or sore throat. These symptoms, based on the definition of ILI, were included. Additionally, we included influenza symptom definitions used by the Centers for Disease Control and Prevention (CDC) [18] and a consultative committee of medical doctors; this second phase generated 29 candidate queries of single words or word combinations consisting of seed keywords and associated words in reference to chief concerns relating to influenza.

### *Using Web Query Recommendations*

Internet search users often require multiple iterations of query refinement to find the desired results from a search engine [16]. Users of search engines can improve their Web search through the help of query recommendations that suggest lists of related queries, allowing users to improve the usability of Web search engines and to access queries that better represent their search intent [17]. We considered queries suggested by keyword recommendations from the Korean websites Daum and Naver In this third phase, entering Flu into the search engines allowed us to identify 75 related queries in the form of single words or word combinations.

## Feature Selection and Prediction Model

We divided the data into training and validation sets. Data from April 3, 2011 to June 29, 2013, were used as the training set for modeling and data from June 30, 2013 to April 5, 2014, were used as the validation set for the model test. Volumes of six seed queries and 146 related queries, obtained after duplicate queries were eliminated from the set of 216 candidate queries, were used for analysis. Before applying the algorithm to each dataset, all data were preprocessed by appropriate transformation and normalization methods. To identify optimal predictors, we applied a least absolute shrinkage and selection operator (Lasso) algorithm. Feature selection can be used to avoid overfitting of irrelevant features and to improve predictive performance (ie, resulting in more rapid and cost-effective predictions) [19,20]. The least absolute shrinkage and selection operator (Lasso) algorithm benefits from a tendency to assign zero weights to irrelevant or redundant features and, hence, is an effective technique for shrinkage and feature selection [21]. Because we



XSL•FO
RenderX

aimed to identify predictors of influenza epidemics, feature selection processing was performed at three time points (defined as lag -2, -1, and 0) on the training set portion of the influenza surveillance data using 10-fold cross-validation. We considered all optimal features selected in each lag for model building.

Support vector machine for regression (SVR) was conducted to construct a model predicting influenza epidemics with selected features. Support vector machines, which are represented as one of the kernel-based methods in supervised machine learning, have been applied successfully to classification tasks and, more recently, also to regression [22]. Grid search and 10-fold cross-validation were performed to select the optimal SVR parameter settings, including the penalty parameter *C* and the kernel function parameter such as the gamma for the radial basis function kernel. Ranges of values for grid search can be summarized as follows (elements in each list denote the beginning, end, and number of samples to generate, respectively): penalty parameter *C* (0.01, 10, 0.01); gamma (0.0001, 1, 0.0001). We assessed the root mean square error (RMSE), particular log errors, and the correlation between predicted values and influenza surveillance data using the validation set. All statistical analyses were performed using the R software package (version 3.0.3; R Development Core Team, Auckland, New Zealand).

### Ethics Statement

This study was exempted from ethical review by the Institutional Review Board of Seoul National University.

## Results

A total of 146 queries related to influenza were generated through our initial query selection approach (see Multimedia Appendix 1). Feature selection was performed based on 152 queries including six seed keywords, and optimal features for the prediction of influenza incidence were chosen using 10-fold cross-validation. Table 1 presents the results of feature selection based on ILI surveillance data. Of the 152 queries, 15, 14, and 29 principal features (the total number of features without duplication=36) exhibited the minimum lambda value in lag-2, lag-1, and 0, respectively. The optimal features for the prediction of ILI incidence were derived from queries with reference to social media data (29/36 features), query recommendations (24/36 features), chief concerns relating to influenza (4/36 features), and seed keywords (1/36 features) (Table 1).

We evaluated the performance of the prediction model, created on the basis of the training set for ILI surveillance, with the validation set. Our results indicated that the SVR model (*C*= 1.32; gamma=0.0002) performed well; the prediction values were highly correlated with recently observed ILI incidence rates ($r$=.956; $P$<.001) (see Figure 1,Multimedia Appendix 2 and Multimedia Appendix 4).

We adopted the same principle with regard to the prediction of virological surveillance as we did with ILI. Table 2 presents the results of feature selection based on virological surveillance data. Of the 152 queries, 28, 26, and 45 principal features (the total number of features without duplication=53) exhibited the minimum lambda value in lag-2, lag-1, and 0, respectively. The optimal features for the prediction of virological incidence were also derived from queries with reference to social media data (42/53), query recommendations (31/53), chief concerns relating to influenza (7/53), and seed keywords (1/53) (Table 2).

Figure 2 shows the result of the performance of the prediction model for virological surveillance. The SVR model (*C*= 2.14; gamma=0.0006) performed well; the prediction values were highly correlated with recently observed virological incidence rates ($r$=.963; $P$<.001) (see Figure 2,Multimedia Appendix 3, and Multimedia Appendix 4).



XSL•FO
RenderX

**Table 1.** Optimal features for influenza-like illness surveillance.

| Query | Query reference | Coefficient | | |
|---|---|---|---|---|
| | | Lag-2 | Lag-1 | Lag 0 |
| (Intercept) | | 0.332 | 0.321 | 0.497 |
| A hyeong influenza [influenza A type] | Social media; query recommendation | 0.745 | 0 | 0.109 |
| A hyeong dokgam [influenza A type] | Social media; query recommendation | 4.928 | 20.154 | 21.503 |
| A hyeong inpeulruenja [influenza A type] | Social media; query recommendation | 0.065 | 0.761 | 1.127 |
| B hyeong influenza [influenza B type] | Social media query; recommendation | 0 | 0 | 0.345 |
| B hyeong dokgam [influenza B type] | Social media; query recommendation | 0 | 0.029 | 1.447 |
| Influenza A | Social media; query recommendation | 2.345 | 0.086 | 0 |
| Influenza A hyeong [influenza A type] | Social media; query recommendation | 1.894 | 0.927 | 0.029 |
| Vaccine | Social media | 0 | 0 | –0.1151 |
| Geongang [health] | Social media | 0.393 | 0.395 | 0.109 |
| Dokgamgamyeom [flu infection] | Social media | 0.052 | 0 | 0 |
| Dokgamgeomsa [flu check] | Social media; query recommendation | 4.303 | 8.893 | 4.402 |
| Dokgam gyeokrigigan [flu isolation period] | Query recommendation | 0 | 0 | 0.177 |
| Dokgam gichim [flu cough] | Social media; chief concern | 0 | 0 | 1.106 |
| Dokgam baireoseu [flu virus] | Social media; query recommendation | 0 | 0 | –0.220 |
| Dokgam yeol [flu fever] | Chief concern | 0.391 | 0 | 0 |
| Dokgam yebang [flu prevention] | Social media; query recommendation | 0 | 0 | –0.152 |
| Dokgam yebangjeopjong [flu vaccination] | Social media; query recommendation | 0 | 0 | –0.1174 |
| Dokgam ipwon [flu hospitalization] | Social media; query recommendation | 0 | 0 | 1.470 |
| Dokgam jeonyeom [flu infection] | Social media; query recommendation | 0 | 0 | 2.569 |
| Dokgam jeonpa [flu dissemination] | Social media; query recommendation | 0.547 | 0.322 | 0.017 |
| Dokgam pyeryeom [flu pneumonia] | Social media ; chief concern | 0 | 0 | 0.005 |
| Dokgam hakgyo [flu school] | Social media | 0 | 0.122 | 0 |
| Dokgam hwanja [flu patient] | Social media | 0.066 | 0 | 0 |
| Soa dokgamjeungsang [child flu symptoms] | Query recommendation | 0.811 | 0.323 | 0.135 |
| Sinjongpeulru jeungsang [new flu symptoms] | Social media; query recommendation | 55.980 | 46.156 | 58.415 |
| Simhangamgi [severe cold] | Social media | 0 | 0 | 0.031 |
| Eorini dokgamyuhaeng [child flu epidemic] | Query recommendation | 0 | 0 | 0.002 |
| Onmomi apeum [whole body pain] | Chief concern | 0 | 0.038 | 0.072 |
| Inpeulruenja geomsa [influenza check] | Social media; query recommendation | 0 | 0.233 | 0 |
| Inpeulruenja yak [influenza medicine] | Social media; query recommendation | 0 | 0 | –0.005 |
| Inpeulruenja yuhaeng [influenza epidemic] | Social media | 0 | 0 | 0.003 |
| Inpeulruenja jeungsang [influenza symptoms] | Social media; query recommendation | 6.254 | 0 | 0 |
| Inpeulruenja jeungse [influenza symptoms] | Social media; query recommendation | 0 | 0 | 0.209 |
| Junggukdokgam [China influenza] | Query recommendation | 0 | 0 | –0.056 |
| Tamipeulru [Tamiflu] | Social media: query recommendation | 0 | 0 | 0.517 |
| Peulru [flu] | Seed keyword | 0.621 | 0.562 | 0.339 |



XSL•FO
RenderX

**Table 2.** Optimal features for virological surveillance.

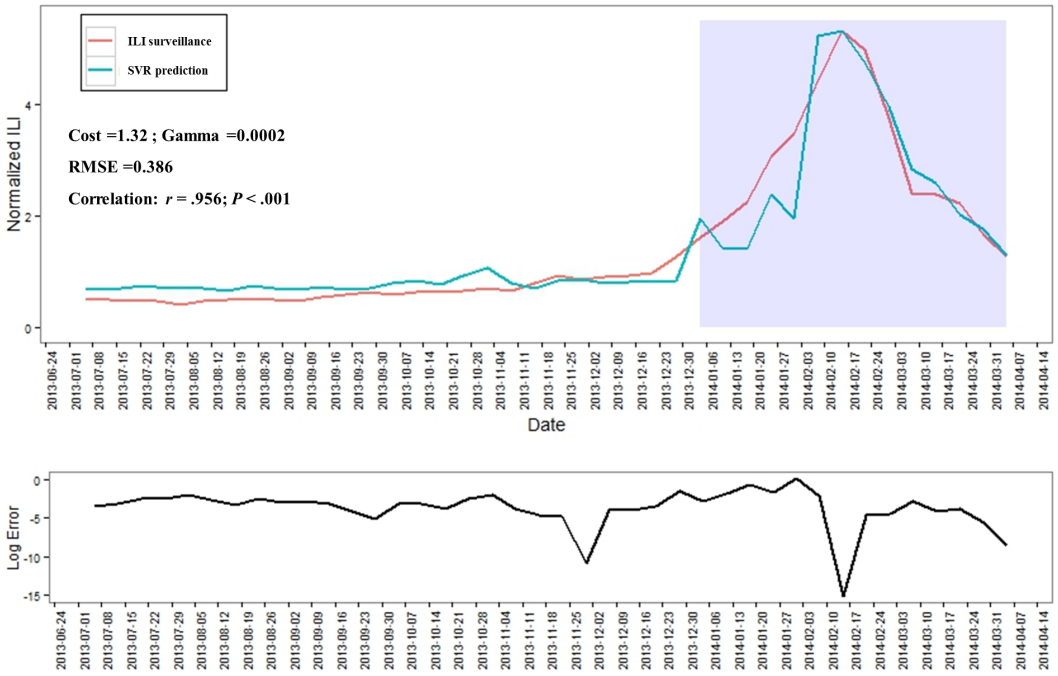
| Query | Query reference | Coefficient | | |
|---|---|---|---|---|
| | | Lag-2 | Lag-1 | Lag 0 |
| (Intercept) | | –1.459 | –3.124 | –2.147 |
| A hyeong influenza [influenza A type] | Social media; query recommendation | 26.413 | 18.899 | 22.579 |
| A hyeong dokgam [ influenza A type] | Social media; query recommendation | 0 | 0 | 379.041 |
| B hyeong dokgam [ influenza B type] | Social media; query recommendation | 6.007 | 15.324 | 24.039 |
| B hyeong dokgamjeungsang [ symptoms of influenza B type] | Social media; query recommendation | 0 | 0 | 0.229 |
| Influenza A | Social media; query recommendation | 37.953 | 25.021 | 17.449 |
| Influenza ahyeong [influenza A type] | Social media; query recommendation | 24.114 | 19.342 | 11.426 |
| Gamgibaireoseu [cold virus] | Social media | 0 | 0 | 4.898 |
| Gamgi pparrinatneunbeop [how to cure flu quickly] | Query recommendation | 5.365 | 4.262 | 2.343 |
| Gamgiyebang [cold prevention] | Social media; query recommendation | 0 | 0 | –0.450 |
| Gamgiyebangbeop [how to prevent a cold] | Social media | –0.155 | –2.736 | –4.140 |
| Geongang [health] | Social media | 4.091 | 3.562 | 3.390 |
| Geunyuktong [muscle pain] | Social media; chief concern | 0 | 0 | –0.265 |
| Nalssi [weather] | Social media | 0 | 0 | –0.111 |
| Dokgam ahyeong [flu A type] | Social media; query recommendation | 0 | 0 | 22.772 |
| Dokgamgamyeom [flu infection] | Social media | 12.236 | 1.449 | 0 |
| Dokgamgeomsa [flu check] | Social media; query recommendation | 38.254 | 31.878 | 0 |
| Dokgam gyeokrigigan [flu isolation period] | Query recommendation | 0 | 0 | 12.145 |
| Dokgam goyeol [flu high fever] | Social media; chief concern | 0 | 0 | 1.745 |
| Dokgam gichim [flu cough] | Social media; chief concern | 0 | 0 | 25.911 |
| Dokgam noin [flu in the elderly] | Social media | 0 | 0 | –3.739 |
| Dokgam baireoseu [flu virus] | Social media; query recommendation | 0 | 0 | –0.777 |
| Dokgam i [flu child] | Social media | 0 | 0 | 2.694 |
| Dokgam eorini [flu child] | Social media | 0 | 0 | –0.477 |
| Dokgam yebang [flu prevention] | Social media; query recommendation | –2.467 | –9.760 | –12.191 |
| Dokgam yebanghaneunbangbeop [how to prevent flu] | Query recommendation | 0 | 0 | –0.638 |
| Dokgam yuhaeng [flu epidemic] | Social media; query recommendation | 0 | 0 | –0.109 |
| Dokgam ipwon [flu hospitalization] | Social media; query recommendation | 8.156 | 0 | 13.793 |
| Dokgam jeonyeom [flu infection] | Social media; query recommendation | 38.184 | 81.830 | 9.762 |
| Dokgam jeonpa [flu dissemination] | Social media; query recommendation | 2.596 | 5.613 | 3.973 |
| Dokgamjusa [flu injection] | Social media; query recommendation | –3.907 | 0 | 0 |
| Dokgamjuuibo [flu watch] | Query recommendation | 0.883 | 0.310 | 0 |
| Dokgam hakgyo [flu school] | Social media | 9.268 | 0 | 0 |
| Dokgam hapbyeongjeung [flu complication] | Social media | 0 | 0 | 3.513 |
| Dokgamhwanja [flu patient] | Social media | 7.024 | 5.027 | 3.205 |
| Dwaejidokgam [swine flu] | Query recommendation | 0.358 | 0 | 0 |
| Maseukeu [mask] | Social media | 8.053 | 0 | 0 |
| Momsal [body aches] | Social media; chief concern | 0 | 1.387 | 3.912 |
| Soa dokgam jeungsang [child flu symptoms] | Query recommendation | 4.737 | 8.058 | 9.041 |
| Adong dokgam jeungsang [child flu epidemic] | Social media; query recommendation | 0 | 0 | –5.273 |



XSL•FO
RenderX

| Query | Query reference | Coefficient | | |
|---|---|---|---|---|
| | | Lag-2 | Lag-1 | Lag 0 |
| (Intercept) | | –1.459 | –3.124 | –2.147 |
| Eoreun dokgam jeungsang [adult flu symptoms] | Query recommendation | 5.156 | 1.485 | 0.610 |
| Eolgultongjeung [face pain] | Chief concern | –1.057 | 0 | 0 |
| Onmomi apeum [whole body pain] | Chief concern | 2.962 | 3.725 | 4.791 |
| Uisa [doctor] | Social media | –3.153 | –0.436 | –0.712 |
| inpeulruenja ahyeong [influenza A type] | Social media; query recommendation | 0 | 8.349 | 5.837 |
| Inpeulruenja samang [influenza death] | Social media; query recommendation | 0 | –0.363 | –5.193 |
| Inpeulruenja yak [influenza medicine] | Social media; query recommendation | 0 | 0 | –0.560 |
| Inpeulruenja jeungse [influenza symptoms] | Social media; query recommendation | 3.039 | 2.051 | 5.303 |
| Ipwon [hospitalization] | Social media | 0 | 0 | –0.213 |
| Joryudokgam [avian flu] | Query recommendation | 3.972 | 4.239 | 3.492 |
| Tamipeulru [Tamiflu] | Social media; query recommendation | 0 | 65.618 | 75.462 |
| Pyeryeom [pneumonia] | Social media; query recommendation; chief concern | 0 | 0 | –1.288 |
| Peulru [flu] | Seed keyword | 15.992 | 13.406 | 5.924 |
| Hwanja [patient] | Social media | –4.543 | –3.170 | –2.922 |

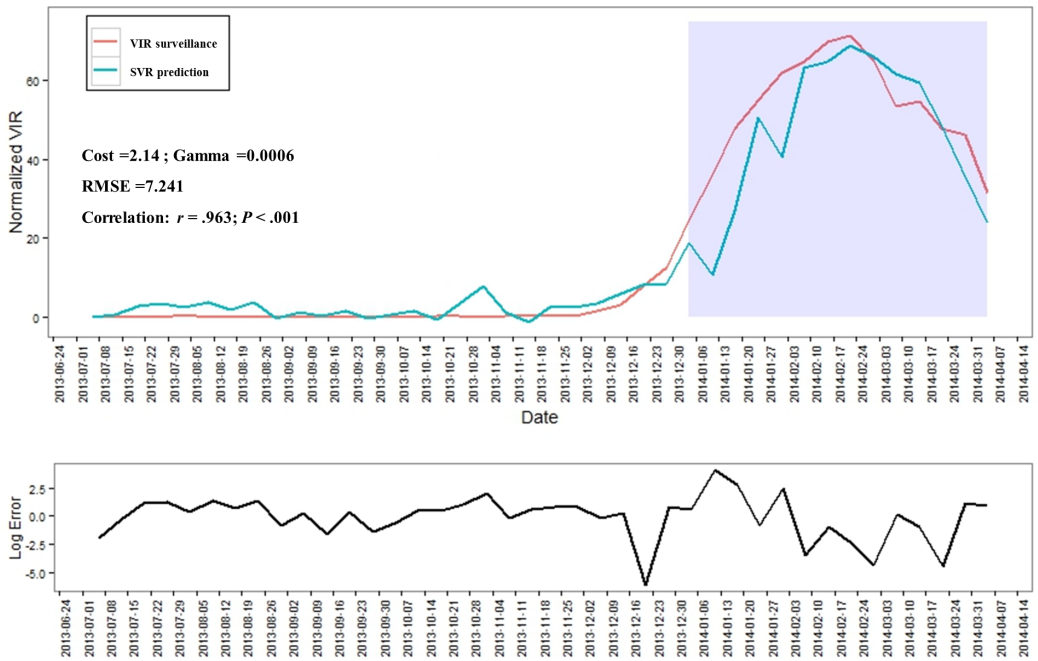
**Figure 1.** Support vector machine for regression(SVR) prediction and error for influenza-like illness (ILI) surveillance in Korea. This figure shows the performance of the SVR model using the validation set of KCDC surveillance data to predict the next observation. Note: log error=log([obs–exp]2/abs[exp]).






XSL•FO
RenderX

**Figure 2.** Support vector machine for regression (SVR) prediction and error for virological surveillance in Korea. This figure shows the performance of the SVR model using the validation set of KCDC surveillance data to predict the next observation. Note: log error=log([obs–exp]2/abs[exp]); VIR: virological positive rate.




## Discussion

This study investigated whether search queries have the capacity to enhance the traditional influenza surveillance system in South Korea. To select queries most likely to be associated with influenza epidemics, we adopted an approach that explored contextual information available in social media data. A considerable proportion of optimal features for our final models were derived from queries with reference to the social media data. Our best model for South Korean ILI data included 36 queries and was highly correlated with observed ILI incidence rates. Our model for virological data, which included 53 queries generated through the same principles as the ILI model, performed equally well in terms of its correlation with observed virological incidence rates. Hence, our models for detecting national influenza incidence have the power to monitor changes. These results demonstrate the feasibility of search queries in enhancing influenza surveillance in South Korea.

Created to predict the incidence of influenza throughout the year, including during high- and low-incidence seasons, our model performed as well as previous models that had benefited from full access to search logs to predict influenza incidence using search queries [3,4,6]. Researchers who do not have full access to search logs need to choose the most pertinent queries, but these may be difficult to determine [1]. Our current approach for query selection using social media data appears to be ideal for supporting influenza surveillance based on search query data. First, it may be helpful for obtaining information for query selection because they contain a greater variety of contextual health information, with diverse descriptions of health states. Above all, it may be a more efficient and unobtrusive way to gather health information. Second, an approach using social media data offers clues for understanding such predictors and their weight, which may vary over time. In generating a prediction model using search query data, it is important to note that search queries change over time. An individual's search behavior changes constantly and keywords submitted by individuals may be influenced by numerous factors, such as media-driven interest or various events [5,23,24]. These changes alter or degrade the performance of search query-based surveillance. The recent Google Flu Trends overestimation can also be understood in the same context [7,8]. Constructing a model that is flexible over time is probably the most difficult, but also the most important, task to complete in the future creation of robust surveillance systems. The systematic exploration of changing predictors in social media data may help to update models based on search queries within a statistical learning framework.

Internet usage is strongly associated with behaviors related to health information seeking and sharing. Some users write expositions about their health through various social media channels, such as blogs and Twitter, while some users leave query logs of health-related questions on the Internet search engines of websites. These types of activities may provide complementary information; it is likely that social media data contain diverse descriptions of personal experiences and information, whereas search engine query data specifically relate to queries, which are submitted for the sole purpose of obtaining information. Starting with studies that have exploited search trends, suggested first in 2006 [25], the notion of detecting influenza activity using Internet-based data has been extended to experimentation with social media data [25]. Thus far, several studies have tried to separately evaluate the scientific potential of each type of novel data for detecting emerging influenza incidence. Although previous empirical studies have reported some significant results, this domain of inquiry is still very much in its infancy [5,23,24] and several limitations pertaining



XSL•FO
RenderX

to data sources can be identified [7,8]. Beyond simply conducting experiments to replicate the findings of previous studies using each type of novel data, perhaps it is time to consider a new strategy, one that adopts mutually reinforcing measures of the valuable information contained in each type of data.

We have used query data obtained from Daum, a Korean local website. The market share of Daum is only 17.4% despite being the second-largest search engine in South Korea; nevertheless, our prediction exhibited strong congruence with national ILI incidence rates. Previous research using query data from Daum has found that some cumulative queries selected by means of survey were also strongly correlated with national influenza surveillance data in South Korea between September 6, 2009 and September 1, 2012 [9]. The findings jointly suggest the possibility of developing an influenza surveillance system using a nondominant search engine.

However, changes in Internet usage rates and health information seeking rates may constitute a somewhat central limitation on the use of search query data. Noise from irrelevant information and uncertainty regarding the representativeness of the sample of health information seekers are also significant limitations. These limitations exist in the data used in our study; thus, optimal features of our model may need to be updated over time.

The initial days of an epidemic represent a critical period for health authorities in terms of initiating appropriate interventions. An online surveillance system allows for cost-effective and near real-time monitoring of infectious disease outbreaks through rapid data collection.

Despite several limitations, this study provides further evidence, based on a new approach, for linkages between the use of Internet-based data and the surveillance of emerging influenza incidence in South Korea. We found that Internet-based influenza surveillance that combines search engine query data with social media data has the power to detect influenza outbreaks, exhibiting strong congruence with traditional surveillance data. Such an approach may provide valuable support in preparing for severe pandemics, such as the 2009 influenza A (H1N1) pandemic, and in controlling seasonal influenza epidemics. Furthermore, in an attempt to exploit the complementary nature of two types of data sources, in this study we fused information drawn from social media with a methodology for query-based influenza surveillance. Our results imply that these new data sources can be compatible and complementary in predicting influenza incidence. Our approach indicates that an online surveillance system can play a significant role in detecting infectious diseases such as influenza in near real time before the release of official reports in South Korea.

## Acknowledgments

This study was supported by the Brain Fusion Research Grant from Seoul National University. The Funder had no role in the study design, data collection and analysis, or preparation of the manuscript. The opinions, results, and conclusions reported in this paper are those of the authors and are independent of the funding source.

## Authors' Contributions

Conceived and designed the experiment: H Woo, Y Cho, E Shim; performed the experiment: Y Cho, H Woo, E Shim, J Lee, C Lee; analyzed the data: H Woo; contributed materials: S Kim; wrote the paper: H Woo, Y Cho.

## Conflicts of Interest

None declared.

## Multimedia Appendix 1

Queries related to influenza generated by an initial query selection approach.

[PDF File (Adobe PDF File), 100KB - jmir_v18i7e177_app1.pdf ]

## Multimedia Appendix 2

Support vector machine for regression(SVR) prediction and error for influenza-like illness(ILI) surveillance in Korea.

[PNG File, 799KB - jmir_v18i7e177_app2.png ]

## Multimedia Appendix 3

Support vector machine for regression(SVR) prediction and error for virological surveillance in Korea.

[PNG File, 803KB - jmir_v18i7e177_app3.png ]

## Multimedia Appendix 4

Search for the optimal final model.



XSL•FO
RenderX

[PDF File (Adobe PDF File), 68KB - jmir_v18i7e177_app4.pdf ]

## References


1. Cho S, Sohn CH, Jo MW, Shin S, Lee JH, Ryoo SM, et al. Correlation between national influenza surveillance data and google trends in South Korea. PLoS One 2013;8(12):e81422 [FREE Full text] [doi: 10.1371/journal.pone.0081422] [Medline: 24339927]
2. Cook S, Conrad C, Fowlkes AL, Mohebbi MH. Assessing Google flu trends performance in the United States during the 2009 influenza virus A (H1N1) pandemic. PLoS One 2011 Aug;6(8):e23610 [FREE Full text] [doi: 10.1371/journal.pone.0023610] [Medline: 21886802]
3. 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]
4. 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]
5. Yuan Q, Nsoesie EO, Lv B, Peng G, Chunara R, Brownstein JS. Monitoring influenza epidemics in china with search query from baidu. PLoS One 2013 May;8(5):e64323 [FREE Full text] [doi: 10.1371/journal.pone.0064323] [Medline: 23750192]
6. Hulth A, Rydevik G, Linde A. Web queries as a source for syndromic surveillance. PLoS One 2009;4(2):e4378 [FREE Full text] [doi: 10.1371/journal.pone.0004378] [Medline: 19197389]
7. Lazer D, Kennedy R, King G, Vespignani A. Big data. The parable of Google Flu: traps in big data analysis. Science 2014 Mar 14;343(6176):1203-1205. [doi: 10.1126/science.1248506] [Medline: 24626916]
8. Lazer D, Kennedy R, King G, Vespignani A. Twitter: big data opportunities--response. Science 2014;345(6193):148-149. [Medline: 25013053]
9. Seo DW, Jo MW, Sohn CH, Shin SY, Lee JH, Yu M, et al. Cumulative query method for influenza surveillance using search engine data. J Med Internet Res 2014 Dec;16(12):e289 [FREE Full text] [doi: 10.2196/jmir.3680] [Medline: 25517353]
10. Kang M, Zhong H, He J, Rutherford S, Yang F. Using Google Trends for influenza surveillance in South China. PLoS One 2013 Jan;8(1):e55205 [FREE Full text] [doi: 10.1371/journal.pone.0055205] [Medline: 23372837]
11. The World in 2015: ICT Facts and Figures. Geneva, Switwerland: International Telecommunication Union; 2014. URL: http://www.itu.int/en/ITU-D/Statistics/Documents/facts/ICTFactsFigures2015.pdf [accessed 2015-07-09] [WebCite Cache ID 6Zt1Nb6VZ]
12. Infectious Disease Statistics System, Korea Centers for Disease Control and Prevention. URL: http://is.cdc.go.kr/dstat/index.jsp [accessed 2015-07-09] [WebCite Cache ID 6Zt0Q7gKi]
13. NAVER. URL: http://section.blog.naver.com [accessed 2015-07-09] [WebCite Cache ID 6Zt0XhwPJ]
14. Daum. URL: http://www.daum.net/ [accessed 2015-07-09] [WebCite Cache ID 6Zt0cdyl1]
15. Internet Trend. 2015. Market share of search engine in South Korea URL: http://www.internettrend.co.kr/trendForward.tsp [accessed 2014-07-06] [WebCite Cache ID 6QrnPuRwJ]
16. He Q, Jiang D, Liao Z. Web query recommendation via sequential query prediction. 2009 Mar 29 Presented at: IEEE 25th International Conference on Data Engineering; Mar 29-Apr 2, 2009; Shanghai p. 1443-1454 URL: http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=4812545 [WebCite Cache ID 6iPJPmh3V]
17. Baeza-Yates R, Hurtado C, Mendoza M. Query recommendation using query logs in search engines. In: Current Trends in Database Technology - EDBT 2004 Workshops. Berlin: Springer Berlin Heidelberg; Mar 14, 2004:588-596.
18. Centers for Disease Control and Prevention. 2015. Flu symptoms and complications URL: http://www.cdc.gov/FLU/ABOUT/disease/symptoms.htm [accessed 2015-07-09] [WebCite Cache ID 6Zt0j4vyV]
19. Guyon I, Elisseeff A. An introduction to variable and feature selection. J Mach Learn Res 2003 Mar 01;3:1157-1182.
20. Saeys Y, Inza I, Larrañaga P. A review of feature selection techniques in bioinformatics. Bioinformatics 2007 Oct 1;23(19):2507-2517 [FREE Full text] [doi: 10.1093/bioinformatics/btm344] [Medline: 17720704]
21. Li F, Yang Y, Xing E. From lasso regression to feature vector machine. 2005 Presented at: Advances in Neural Information Processing Systems 18; Dec 5-8, 2005; Vancouver, BC p. 411-418.
22. Smola AJ, Schölkopf B. A tutorial on support vector regression. Stat Comput 2004 Aug;14(3):199-222. [doi: 10.1023/B:STCO.0000035301.49549.88]
23. Milinovich GJ, Williams GM, Clements AC, Hu W. Internet-based surveillance systems for monitoring emerging infectious diseases. Lancet Infect Dis 2014 Feb;14(2):160-168. [doi: 10.1016/S1473-3099(13)70244-5] [Medline: 24290841]
24. Althouse BM, Ng YY, Cummings DA. Prediction of dengue incidence using search query surveillance. PLoS Negl Trop Dis 2011 Aug;5(8):e1258 [FREE Full text] [doi: 10.1371/journal.pntd.0001258] [Medline: 21829744]
25. Bernardo TM, Rajic A, Young I, Robiadek K, Pham MT, Funk JA. Scoping review on search queries and social media for disease surveillance: a chronology of innovation. J Med Internet Res 2013;15(7):e147 [FREE Full text] [doi: 10.2196/jmir.2740] [Medline: 23896182]


## Abbreviations

**API:** application program interface



XSL•FO
RenderX

**ILI:** influenza-like illness
**KCDC:** Korea Center for Disease Control and Prevention
**Lasso:** least absolute shrinkage and selection operator
**RMSE:** root mean square error
**SVR:** support vector machine for regression

*Edited by G Eysenbach; submitted 25.07.15; peer-reviewed by E Nsoesie, SY Shin; comments to author 03.09.15; revised version received 17.04.16; accepted 19.05.16; published 04.07.16.*

*Please cite as:*
*Woo H, Cho Y, Shim E, Lee JK, Lee CG, Kim SH*
*Estimating Influenza Outbreaks Using Both Search Engine Query Data and Social Media Data in South Korea*
*J Med Internet Res 2016;18(7):e177*
*URL: http://www.jmir.org/2016/7/e177/*
*doi:10.2196/jmir.4955*
*PMID:27377323*

©Hyekyung Woo, Youngtae Cho, Eunyoung Shim, Jong-Koo Lee, Chang-Gun Lee, Seong Hwan Kim. Originally published in the Journal of Medical Internet Research (http://www.jmir.org), 04.07.2016. This is an open-access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/2.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on http://www.jmir.org/, as well as this copyright and license information must be included.



XSL•FO
RenderX

Original Paper

# Health Care Applicability of a Patient-Centric Web Portal for Patients' Medication Experience

Song Hee Hong[1], PhD; Woojung Lee[1], PharmD; Yazed AlRuthia[2], PharmD, PhD

[1]College of Pharmacy, Seoul National University, Seoul, Republic Of Korea

[2]College of Pharmacy, King Saud University, Riyadh, Saudi Arabia

**Corresponding Author:**
Song Hee Hong, PhD
College of Pharmacy
Seoul National University
College of Pharmacy Bldg 20-210
Seoul,
Republic Of Korea
Phone: 82 2 880 1547
Fax: 82 2 877 7859
Email: songhhong@snu.ac.kr

## Abstract

**Background:** With the advent of the patient-centered care paradigm, it is important to examine what patients' reports of medication experience (PROME) mean to patient care. PROME available through a Web portal provide information on medication treatment options and outcomes from the patient's perspective. Patients who find certain PROME compelling are likely to mention them at their physician visit, triggering a discussion between the patient and the physician. However, no studies have examined PROME's potential applicability to patient care.

**Objective:** This study aimed to examine older (≥50 years) adults' perceptions of the health care applicability of a hypothetical PROME Web portal. Specifically, this study investigated whether PROME would facilitate patient-physician communication, and identified the preferred reporting items and the trusted sponsors of such a PROME Web portal.

**Methods:** We used a cross-sectional, self-administered, 5-point Likert scale survey to examine participants' perceptions of a hypothetical PROME Web portal that compared PROME for 5 common antihypertensive medications. Between August and December 2013, we recruited 300 members of 7 seniors' centers in a metropolitan area of a southeastern state of the United States to participate in the survey.

**Results:** An overwhelming majority of study participants (243/300, 81.0%) had a favorable perception of PROME's health care applicability. They were mostly positive that PROME would facilitate patient-physician communication, except for the perception that physicians would be upset by the mention of PROME (n=133, 44.3%). Further, 85.7% (n=257) of participants considered the PROME information trustworthy, and 72.0% (n=216) were willing to participate by reporting their own medication experiences. Study participants wanted the PROME Web portal to report the number of reviews, star ratings, and individual comments concerning different medication attributes such as side effects (224/809, 27.7%), cost (168/809, 20.8%), and effectiveness (153/809, 18.9%). Finally, the PROME Web portal sponsorship was important to participants, with the most trusted sponsor being academic institutions (120/400, 30.0%).

**Conclusions:** PROME, if well compiled through Web portals, have the potential to facilitate patient-physician communication.

***(J Med Internet Res 2016;18(7):e202)*** doi:10.2196/jmir.5813

**KEYWORDS**

patient-physician communication; medication experience outcomes; patient reports; Internet; patient-centered practice; Web portal



XSL•FO
RenderX

## Introduction

There is now consensus that patients' reports of their health experience reflect quality of care [1]. Accordingly, payers such as the Centers for Medicare and Medicaid Services in the United States and the National Health Service in the United Kingdom use patient-reported experience measures for the purpose of performance evaluation and compensation of health care providers [1-4]. However, no studies have examined what patients' reported medication experience means to patient care. The frequent use of medication indicates that such information could have great potential to affect patient care, especially in the management of chronic diseases.

Patients' reports of medication experience (PROME) are likely to facilitate patient-physician communication. Patients who find certain PROME compelling are likely to mention them at their physician visit. The mention then would trigger a discussion between the patient and the physician, just as direct-to-consumer advertising (DTCA) does. One-third of the participants in a US Food and Drug Administration (FDA) survey of DTCA said that they initiated conversations with their physicians because of advertising [5]. PROME may more effectively trigger patient-physician communication than DTCA because PROME comes from users, while advertising comes from sellers. Increased patient-physician communication is key to advancing patient-centered practice.

The potential for PROME to influence patient care has given birth to several Web portals such as AskaPatient [6] and DrugRatingz [7]. WebMD [8] and ConsumerReports [9] have also begun to compile patient reviews of medications, along with drug information. The Web portals provide a venue for patients to report their medication experience in terms of effectiveness, side effects, and costs. Moreover, PROME Web portals can present information according to medication classes and patient characteristics. Patients who are browsing those PROME Web portals can easily learn about what medication options have received favorable ratings from which group of patients. When patients come across a report from other patients in the same situation, they are likely to act on the information included in the report [10]. These medicine-focused social media, with a large volume of high-quality first-hand patients' reviews, are also considered to be a promising data source for understanding patients' medication experience [11].

With growth in the number of PROME Web portals comes a critical need to examine PROME's potential applicability to health care. In this study, we aimed to determine participants' perceptions of whether PROME would facilitate patient-physician communication, and to identify the preferred patient-reporting items and trusted sponsors of such a PROME Web portal.

## Methods

### Study Design and Participants

We used a cross-sectional survey to examine participants' perceptions of PROME. A detailed description of the survey procedure is given in a doctoral dissertation [12].We scheduled visits to conduct the survey with the coordinators of 7 seniors' centers in a metropolitan area of a southeastern state in the United States. On each visit, 2 research assistants recruited study participants into a reserved private room and explained the purpose of the study, along with the rights of the participants. Those who completed the survey received a US $20 grocery gift card as an appreciation for their participation. Data collection was started in August 2013 and was continued until we reached our goal of 300 completed surveys in December 2013. Before beginning the study, the University of Tennessee Health Science Center Institutional Review Board determined the study to be exempt from their oversight.

### Survey Instrument

The survey instrument used for this study contained a chart from a hypothetical PROME Web portal that compared PROME for 5 common antihypertensive medications (Figure 1). The chart used arbitrary 5-star ratings and included the number of people who supposedly gave reviews for each medication. We asked for the following sociodemographic information from the participants: sex, age, years of education, race, family member(s) they live with, and income. We asked participants 17 questions in total: 6 demographic questions and 11 questions about the PROME Web portal.

Referring to the PROME chart, we surveyed participants' perceptions of its potential applicability to health care using 7 questions: (1) 1 question on overall usefulness, (2) 4 questions on patient-physician communication, and (3) 2 final questions: 1 on the perceived credibility of the information provided by the PROME Web portal, and 1 on the willingness of the survey participants to provide their own medication experiences to a PROME Web portal. The 4 questions on patient-physician communication concerned the likelihood for patients to mention PROME to their physician, the likelihood for PROME to facilitate the communication, the perceived likelihood for physicians to be upset by the PROME mention, and the likelihood for patients to ask their physicians to prescribe the PROME-recommended medication. All the questions were answered on a 5-point Likert scale with the following choices: definitely, very probably, probably, probably not, and definitely not. Previous surveys on the effect of DTCA on patient-physician communication indicated that there would be more positive than negative evaluations [5]. We used the unbalanced scale to provide more discrimination between positive evaluations [13]

The survey also had 2 questions on the preferred reporting items, that is, what information the participants wanted to see in the PROME Web portal. The first question asked participants to indicate which reporting items (star ratings, number of reviews, and individual comments) they believed valuable. The second question asked participants to indicate any medication attributes (effectiveness, side effect, food interaction, convenience, and cost) they believed the PROME Web portal should report.

Lastly, the survey had 2 additional questions on PROME Web portal sponsorship, that is, what entity study participants believed should sponsor the PROME Web portal. The first question asked participants to rate the importance of PROME Web portal sponsorship on a 5-point Likert scale. The second



XSL•FO
RenderX

question asked participants to indicate any types of sponsors (academic institutions, nonprofit foundations, chain pharmacies, health information companies, and drug plans) they would trust.

**Figure 1.** Hypothetical summary of patients' reports of medication experience (PROME) with antihypertensive medications presented to survey participants.




### Statistical Analysis

For the participants' perceptions of PROME's potential applicability to health care, we reported the percentages selected for each of the 5 response choices as a bar graph. Further, we dichotomized the 5 response choices into having a positive (definitely, very probably, and probably) or a negative perception (probably not or definitely not), and then estimated the probability of having a positive perception, along with 95% CIs. We constructed pie charts to describe the frequency distribution of medication attributes that study participants picked as important elements of PROME coverage and to generate the frequency distribution of PROME Web portal sponsors that study participants picked as trusted. We used chi-square tests to determine whether participants' perceptions varied with their sociodemographic characteristics at a significance level of 5%. Analyses were conducted using SAS software, version 9.3 (SAS Institute Inc).

## Results

### Study Participants

We contacted older adults (≥50 years) who were members of seniors' centers in a metropolitan area in a southeastern US state to participate in our survey. Table 1 lists the characteristics of the 300 older adults who completed our survey. They had a mean age of 71.95 years (SD 8.65), were mostly non-Hispanic white (164/300, 54.7%) and female (231/299, 77.3%), with at least some high school education (287/297, 96.6%) and annual incomes of at least US $10,000 (239/262, 92.3%).



XSL•FO
RenderX

**Table 1.** Sociodemographic characteristics of the study participants (N=300).

| Characteristics | No. (%)[a] |
|---|---|
| **Age range (years) (n=295)** | |
| 50–59 | 9 (3.1) |
| 60–69 | 118 (40.0) |
| 70–79 | 105 (35.6) |
| 80–89 | 56 (19.0) |
| ≥90 | 7 (2.4) |
| **Sex (n=299)** | |
| Male | 68 (22.7) |
| Female | 231 (77.3) |
| **Education (n=297)** | |
| Middle school or less | 10 (3.4) |
| High school or graduate | 123 (41.4) |
| Some college | 92 (31.0) |
| College graduate or higher | 72 (24.2) |
| **Race (n=300)** | |
| Non-Hispanic white | 164 (54.7) |
| Non-Hispanic black | 121 (40.3) |
| Other[b] | 15 (5.0) |
| **Living status (n=295)** | |
| Alone | 132 (44.7) |
| With daughter or son | 31 (10.5) |
| With companion or sibling | 21 (7.1) |
| With spouse | 99 (33.6) |
| Other[c] | 12 (4.1) |
| **Income (US$) (n=262)** | |
| <10,000 | 23 (7.7) |
| 10,000–29,000 | 111 (37.0) |
| 30,000–49,000 | 56 (18.7) |
| 50,000–69,000 | 41 (13.7) |
| ≥70,000 | 31 (10.3) |

[a]Some numbers do not add up to 300 because not all participants answered each question.

[b]Other includes Asian, Native American, and Alaskan native.

[c]Other includes living with a parent, a grandson, a niece, or a pet, and living in a retirement community.

## Health Care Applicability of PROME

As Figure 2 shows, an overwhelming majority of study participants (n=243, 81.0%) were positive about the overall usefulness of PROME; however, the percentage positive decreased to 62.3% (187/300) when excluding probably as the response. As for patient-physician communication, 245 of the 300 participants (81.7%) said that they would mention PROME to their physician (166/300, 55.3% for the responses definitely and very probably), and 248 participants (82.9%) said that PROME would facilitate patient-physician communication (138/299, 46.2% for the responses definitely and very probably). However, 133 participants (44.3%) said that their physician would get upset if they mentioned PROME (42/299, 14.1% for the responses definitely and very probably). Further, 209 (69.9%) of study participants were positive that they would ask their physician to prescribe a PROME-recommended medication. When we further broke down the positive responses to the question on whether physicians would be upset by such a request into probably, very probably, and definitely, we found that more participants chose probably (n=110, 36.8%) than very probably and definitely combined (n=99, 33.1%). In other

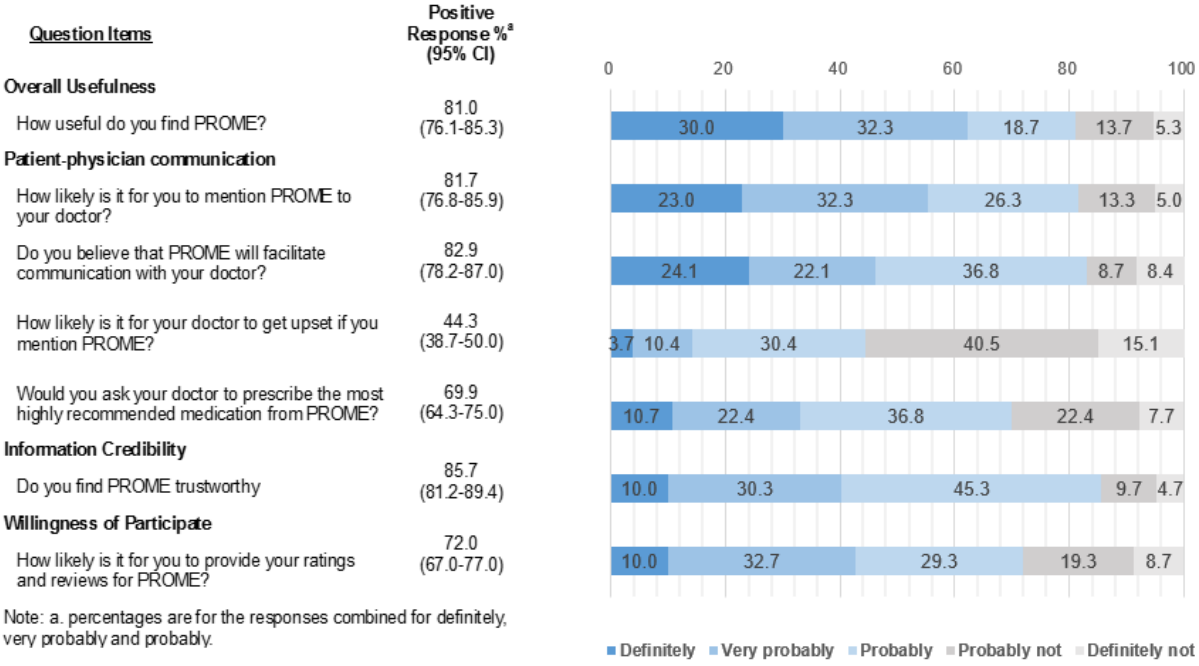


XSL•FO
RenderX

words, study participants had some reservations about asking their physician to prescribe a PROME-recommended medication.

As for information credibility, most of our study participants (n=257, 85.7%) were positive that PROME information is trustworthy. Further, two-thirds (n=216, 72.0%) were willing to participate in PROME Web portals by providing their own medication experiences. However, for information credibility, more participants responded probably (n=136, 45.3%) than very probably and definitely combined (n=121, 40.3%) (Figure 2).

**Figure 2.** Participants' perceptions of the potential health care applicability of patients' reports of medication experience (PROME).




## Preferred Reporting Items and Sponsorship of the PROME Web Portal

When asked about how PROME should be displayed in a Web portal, of the 308 answers given, participants most frequently picked the number of reviews (n=105, 34.1%), followed by star ratings (n=97, 31.5%) and individual comments (n=96, 31.2%) (Figure 3). However, the difference was negligible. Study participants were also asked to indicate medication attributes (such as effectiveness, side effects, ease of use, costs, and interaction with food) that PROME should cover. Study participants gave 809 answers and picked side effects most often (n=224, 27.7%) and ease of use least often (n=110, 13%). Cost (n=168, 20.8%), effectiveness (n=153, 18.9%), and interaction with food (n=151, 18.7%) were picked almost equally (Figure 3). When asked to indicate the importance of the PROME Web portal's sponsorship, 263 of 298 participants (88.2%) said that sponsorship is important (Figure 4). Academic institutions such as the University of Tennessee were viewed as the most trusted sponsors (120/400, 30.0%), followed by nonprofit foundations such as the American Heart Association (97/400, 24.3%) (Figure 4).



XSL•FO
RenderX

**Figure 3.** Left: Reporting items that participants believe are valuable (308 responses). Right: Medication attributes on which participants believe patients' reports of medication experience should report (809 responses).

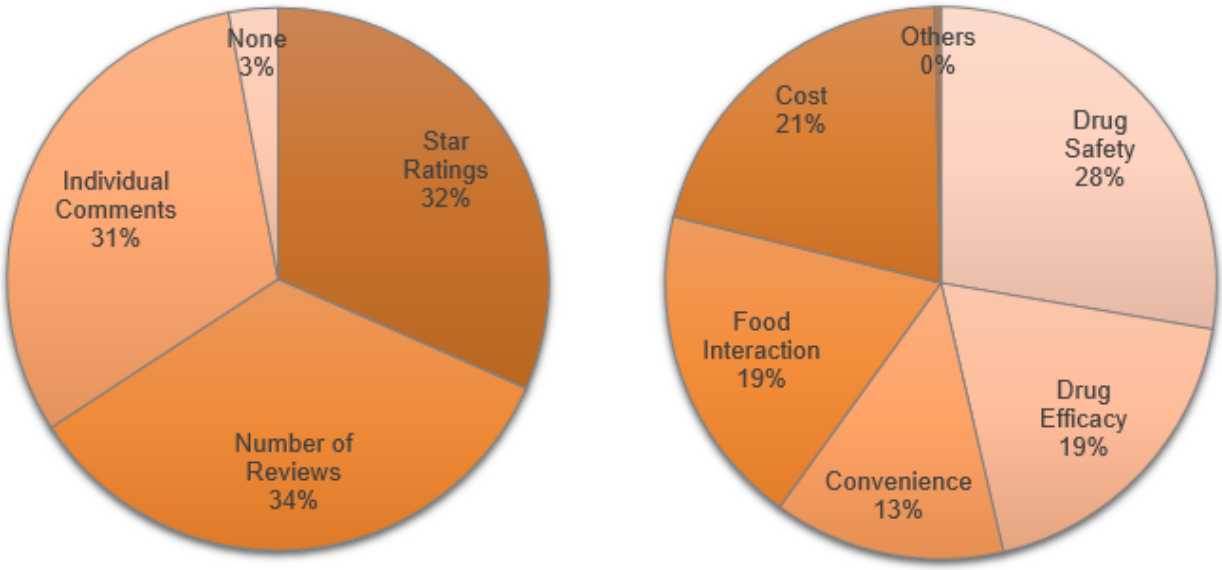
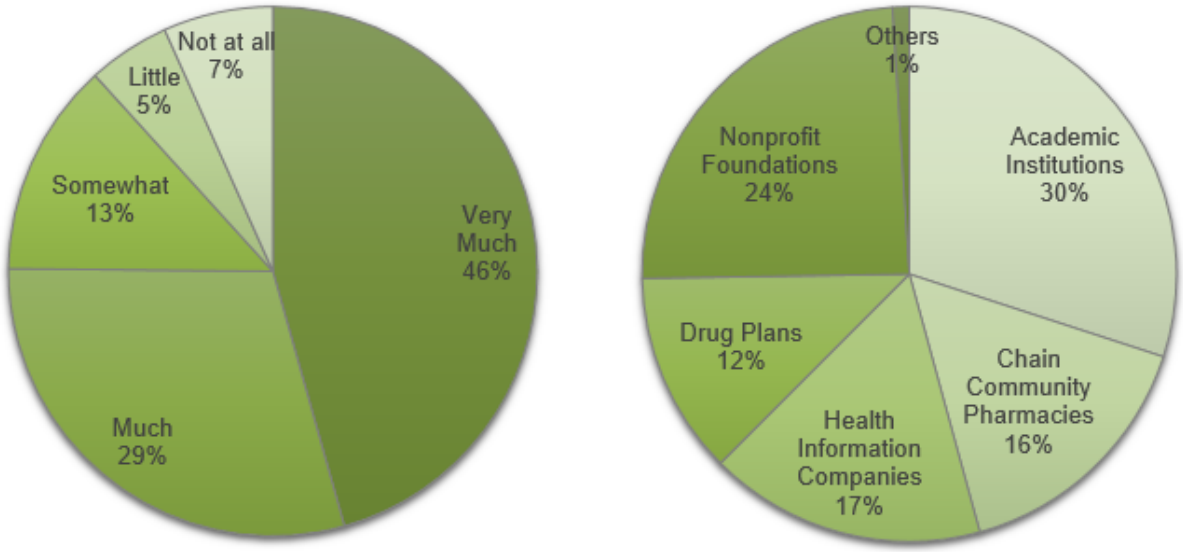



**Figure 4.** Left: The importance of sponsorship of a Web portal featuring patients' reports of medication experience (298 responses). Right: The sponsors that participants trust the most (400 responses).




## Views on PROME According to Demographic Factors

Across all sociodemographic characteristics, study participants were positive that PROME provide useful and credible information to facilitate patient-physician communication (Table 2). Further, their willingness to participate in PROME Web portals remained high across all sociodemographics except for income. Study participants in the highest income bracket (over US $70,000 per year) were least willing to participate in PROME Web portals ($P$=.02) (Figure 5). However, the perceived likelihood that physicians would be upset was significantly different across several demographic variables. With increased education, study participants were more likely to believe that physicians would get upset by a mention of PROME (3/10, 30% for people with middle school education or less and 61/92, 66% for people with some college) (Figure 6). This trend was also present with income levels: the higher the income, the more likely the participant was to believe physicians would get upset by a mention of PROME (Figure 5). Likewise, non-Hispanic whites were more likely than non-Hispanic blacks to believe that physicians would get upset by a mention of PROME.



XSL•FO
RenderX

**Table 2.** Percentage of positive views[a] on the potential applicability of a Web portal featuring patients' reports of medication experience (PROME) to health care, by demographic characteristic (N=300).

| Characteristics | No. | Items relating to perceptions of PROME Web portal applicability, % (95% CI) | | | | |
|---|---|---|---|---|---|---|
| | | Overall usefulness | Patient-physician communication[b] | Physician getting upset | Information credibility | Willingness to participate |
| | | n (%)<br>95% CI | n (%)<br>95% CI | n (%)<br>95% CI | n (%)<br>95% CI | n (%)<br>95% CI |
| **Sex** | | | | | | |
| Female | 231 | 187 (81.0%)<br>75.9–86.0 | 168 (72.7%)<br>66.5–78.4 | 134 (58.0%)<br>51.6–64.4 | 197 (85.3%)<br>80.7–90.0 | 162 (70.1%)<br>64.2–76.1 |
| Male | 68 | 55 (80.9%)<br>71.5–90.2 | 53 (77.9%)<br>66.2–87.1 | 33 (48.5%)<br>36.7–60.4 | 59 (86.8%)<br>78.5–95.0 | 53 (78.0%)<br>67.8–88.1 |
| **Education (years)** | | | | | | |
| ≤8 | 10 | 9 (90.0%)<br>78.0–100.0 | 8 (80.0%)<br>70.1–91.9 | 3 (30.0%)<br>13.1–45.0 | 7 (70.0%)<br>41.6–98.4 | 6 (60.0%)<br>29.6–90.4 |
| 9–12 | 123 | 102 (83.3%)<br>76.1–90.6 | 89 (72.6%)<br>63.9–81.2 | 62 (50.1%)<br>40.9–61.0 | 107 (87.0%)<br>81.1–93.0 | 91 (74.0%)<br>66.2–81.7 |
| 13–16 | 92 | 69 (75.0%)<br>66.2–83.9 | 64 (70.1%)<br>61.4–80.0 | 61 (66.3%)<br>56.6–76.0 | 80 (87.0%)<br>78.3–93.1 | 68 (73.9%)<br>64.9–82.9 |
| ≥17 | 72 | 58 (80.6%)<br>71.4–89.7 | 55 (76.4%)<br>66.6–86.2 | 45 (62.5%)<br>51.3–73.7 | 61 (84.7%)<br>76.4–93.0 | 48 (66.7%)<br>54.6–77.3 |
| **Race** | | | | | | |
| Non-Hispanic white | 164 | 133 (81.1%)<br>75.1–87.1 | 117 (71.3%)<br>64.4–78.3 | 103 (62.8%)<br>55.4–70.2 | 139 (84.8%)<br>78.3–89.9 | 110 (67.1%)<br>59.9–74.3 |
| Non-Hispanic black | 121 | 97 (80.0%)<br>72.8–87.2 | 93 (76.7%)<br>69.1–84.2 | 60 (49.2%)<br>40.2–58.1 | 103 (85.1%)<br>78.9–91.5 | 94 (77.7%)<br>70.3–85.1 |
| Other[c] | 15 | 13 (87.5%)<br>71.3–100.0 | 12 (81.3%)<br>54.4–96.0 | 5 (33.3%)<br>9.5–54.0 | 15 (100.0%)<br>100.0–100.0 | 12 (80.0%)<br>59.8–100.0 |
| **Income (US$)** | | | | | | |
| <10,000 | 23 | 17 (73.9%)<br>56.0–91.7 | 18 (78.3%)<br>61.4–95.1 | 11 (47.8%)<br>27.4–68.2 | 16 (69.6%)<br>50.8–88.4 | 17 (73.9%)<br>56.0–91.9 |
| 10,000–29,000 | 111 | 94 (84.4%)<br>77.6–91.2 | 92 (82.6%)<br>75.5–89.7 | 57 (51.4%)<br>42.0–60.8 | 99 (89.2%)<br>83.4–95.0 | 87 (78.4%)<br>70.7–86.0 |
| 30,000–49,000 | 56 | 47 (84.5%)<br>75.2–93.8 | 41 (74.1%)<br>62.9–85.4 | 31 (55.2%)<br>42.4–68.0 | 50 (87.5%)<br>78.8–96.2 | 44 (78.6%)<br>67.8–89.3 |
| 50,000–69,000 | 41 | 33 (80.5%)<br>68.4–92.6 | 28 (68.3%)<br>54.1–82.5 | 29 (70.7%)<br>54.5–83.9 | 35 (85.4%)<br>74.6–96.2 | 29 (70.7%)<br>56.8–83.9 |
| 70,000 | 31 | 22 (71.0%)<br>55.0–87.0 | 18 (58.1%)<br>40.7–75.4 | 21 (67.8%)<br>48.6–83.3 | 25 (80.7%)<br>62.5–92.6 | 15 (48.4%)<br>30.8–66.0 |

[a]The percentage of positive answers is for the response choices of probably, very probably, and definitely combined.

[b]The 3 items related to patient-physician communication were combined because individually they did not have any significant demographic variations. The percentages of positive responses were for those with the combined score 9, the score when all 3 items were given a rating of 3 (a response of probably on the 5-point scale).

[c]Other includes Asian, Native American, and Alaskan native.



XSL•FO
RenderX

**Figure 5.** Income effects on participants' perceptions of the likelihood that physicians would get upset if they mentioned patients' reports of medication experience (PROME) and their willingness to participate in a PROME Web portal. Error bars indicate 95% CI.

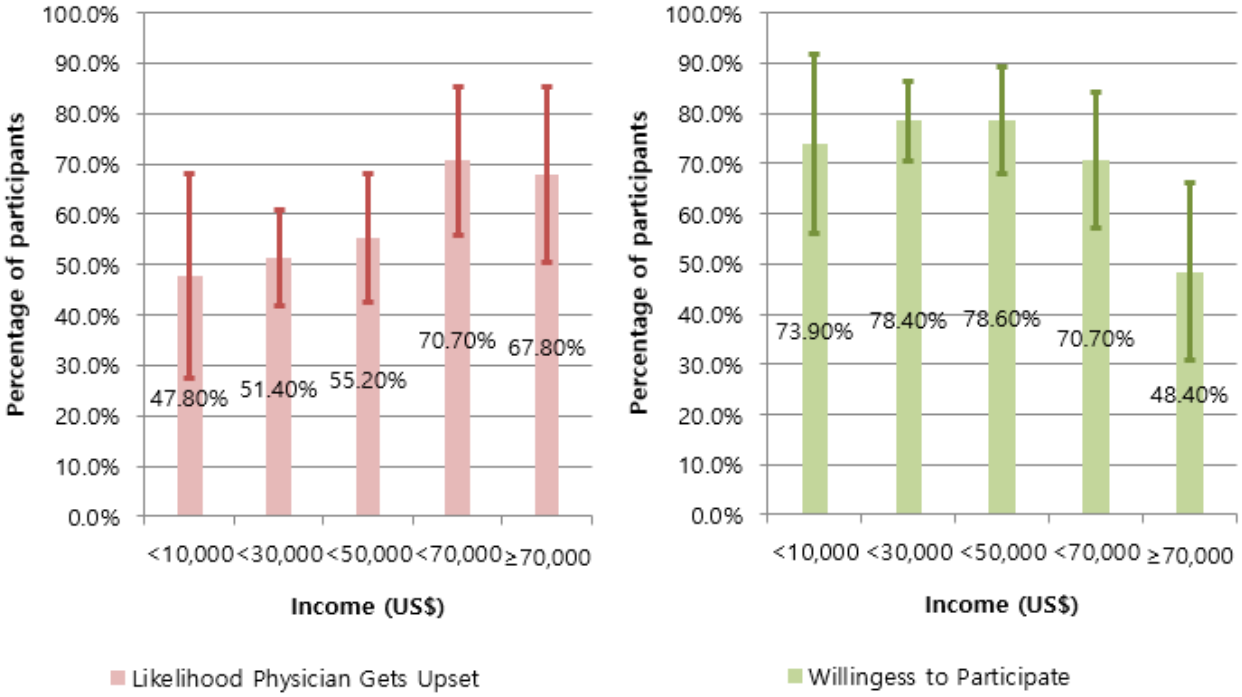
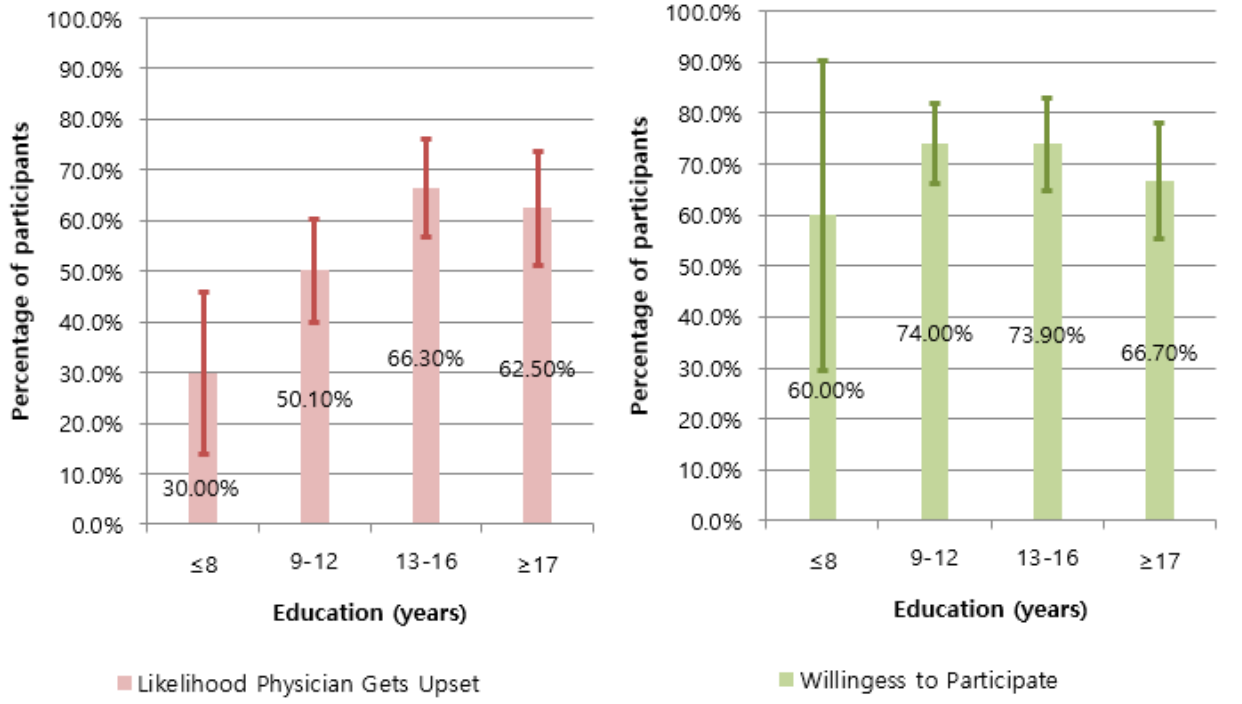



**Figure 6.** Education effects on participants' perceptions of the likelihood that physicians would get upset if they mentioned patients' reports of medication experience (PROME) and their willingness to participate in a PROME Web portal. Error bars indicate 95% CI.




## Discussion

PROME would greatly affect patient care if patients consider the PROME information to be useful and credible, and are willing to contribute their own reviews and ratings for PROME. This study found that an overwhelming majority of study participants (81.0%) were at least positive that PROME are overall useful; 62.3% were definitely or very probably positive. The positive perception implies that older patients are in need of medication information written by the patient, for the patient. Patients who have received treatment under the paternalistic care environment are longing for information about overall treatment processes and alternative treatment options [14].

More than 80% of study participants were positive that PROME would facilitate patient-physician communication by triggering a mention of PROME. This percentage is much higher than



XSL•FO
RenderX

those found in FDA-sponsored surveys of patients regarding DTCA. Only one-third of the participants in the FDA survey conducted in 2002 said that advertisements had encouraged them to ask a question of their physician, while 43%, a decline from 62% in 1999, felt that advertisements helped them have better discussions with their doctors [5]. Evidently, PROME have a greater potential to influence patient-physician communication than advertisements. Further, the likelihood of asking for a specific brand was also influenced quite differently by PROME and advertisements. In our study, more than two-thirds of study participants (69.9%) said that they would ask their physician to prescribe a PROME-recommended brand, compared with 39% of the FDA's survey participants stating that they would ask for an advertisement-recommended brand.

It is interesting that participants perceived that physicians would be upset by their mention of PROME. While our survey did not provide data to explain this perception, perhaps the participants believed that they would be challenging the physician's authority to prescribe by mentioning PROME-recommended drug therapy options. Alternatively, participants might have had past experience with physicians who were reluctant to discuss potential drug therapy options. However, studies have shown that physicians are in fact willing to discuss therapy options with patients [15,16]. It may be, then, that participants felt that physicians, in general, welcome questions on health issues but may not be as welcoming to those on drug therapy options, especially a request for a specific brand.

Patients often regard a piece of information as trustworthy when it comes from other patients with a similar condition [10]. However, study participants had some reservations concerning the trustworthiness of the PROME information. Traditionally, online information has been viewed as less trustworthy than print information [17]. Further, many are reluctant to trust online information [18,19], especially older adults, such as the study participants, who trust online health information less than do younger adults [20,21].

As for willingness to participate in PROME Web portals, 72.0% of study participants said they were willing to provide their own reviews and ratings. It is remarkable that such a high percentage of seniors were willing to participate in PROME Web portals. It may reflect the ongoing trend of a rapidly increasing senior population searching for health-related information on the Internet [22]. It could also reflect the patient-centric health care paradigm, where patients are actively seeking other patients' experience. According to a study of PatientsLikeMe, patients refer to other patients' experience to better understand and control their diseases [10]. Thus, it follows that patients who seek other patients' experience would be willing to participate in PROME Web portals.

Patient reports, just like any other consumer review, can be summarized in 3 reporting items (star ratings, number of reviews, and individual comments). Study participants indicated that the PROME Web portal should provide information on all 3 reporting items. The 3 items evidently capture different aspects of patients' medication experience. Without capturing all these aspects, PROME may not successfully reflect patients' true medication experience.

A well-designed PROME Web portal should also provide information on all the attributes of medication experience, such as effectiveness, side effects, ease of use, costs, and interaction with food. Participants picked side effects, followed closely by cost, effectiveness, and interaction with food, as the most important attributes to report in a PROME Web portal. Ease of use was least often picked. It is not surprising that study participants were most interested in side effects. The FDA's study also reported that far more people look for information on side effects than on benefits (61% vs 10%) [5]. What is rather surprising is that more people in this study wanted the PROME Web portal to report on drug cost than on effectiveness (20.8% vs 18.9%). This finding is starkly different from the FDA's survey, where few people (4%) wanted cost information from advertisements [5]. Patients seem to put more trust in PROME than in advertisements for drug cost information. Chronic diseases are prevalent among older adults and require ongoing medication management. Older patients who live on a fixed income could face substantial financial distress due to drug cost [23,24]. Thus, these patients would naturally seek information on drug cost.

Information available on the Internet has a high chance of misleading people [19,25]. Without a reputable sponsor, online information is difficult to trust [26]. As many as 90% of study participants said that it is important who sponsors the PROME Web portal. Evidently, sponsorship makes a big impact on patients' willingness to trust online medical information. Patients rarely read online information unless there is a transparent and dependable sponsorship [27].Without an address or a phone number of the sponsor, patients simply do not trust online information [28]. As for a trusted sponsor of the PROME Web portal, in this study participants picked academic institutions as the most trusted. Another study also reported that the most trusted sponsor of health information is a university [29]. According to our participants, private organizations such as WebMD, ConsumerReports, and chain pharmacies were viewed as least trusted. People seem to perceive that private organizations act in their own interests ahead of patients' interests.

Across all demographics, most study participants had a favorable view of the health care applicability of PROME. Meanwhile, about one-fifth of study participants did not agree with the usefulness of PROME. Study participants in the highest income bracket (over US $70,000 per year) showed the lowest willingness to participate in PROME Web portals ($P<.05$). The inverse relationship could imply that participants with the highest income are most satisfied with the medical information they have access to and thus are least likely to feel the need for PROME Web portals for additional information. Participants with higher income may have easier access to various medical information resources. Further, the perceived likelihood that physicians would get upset by their patients mentioning PROME was significantly higher among people with greater education and income. Moreover, this perception was more apparent among non-Hispanic whites than among non-Hispanic blacks. It would be interesting to know why participants with more education were more likely to perceive that physicians would



XSL•FO
RenderX

get upset by a mention of PROME. Perhaps education trains people to be more skeptical.

### Limitations

Our findings should be interpreted carefully because of three potential study limitations. First, we have measured participants' perceptions based on convenience sampling in one metropolitan area. Participants' perceptions of the study sample may well have been different from those of the US population as a whole. However, the characteristics of our study participants were similar to the population who were over 50 years of age and lived in a metropolitan area of a southeastern state.

Second, study participants may not have been willing to disclose some information, especially when the information was sensitive because of privacy concerns. For instance, many of the respondents refused to disclose their estimated annual income even though we had guaranteed the anonymity and confidentiality of their responses. While this limitation is common to all survey studies, the study sample could have been more reluctant to release the sensitive information.

Third, this study used an unbalanced 5-point Likert scale (3 positive and 2 negative responses) to elicit study participants' perceptions of PROME with regard to patient-physician communication. Although this scale provides more discrimination between positive responses [13,30], some neutral responses could have been forced to positive ones. When this happens, dichotomizing the responses into positive and negative responses could inflate the occurrence of positive evaluations. However, the likelihood is minimal considering survey participants are known to choose a response based on a label rather than the position on the 5-point scale [31,32]

### Conclusion

This study found that older participants across most demographics considered PROME to provide useful and credible information to facilitate patient-physician communication, and thus were willing to participate in PROME Web portals by sharing their own medication experiences. These participants also believed that an academic institution should sponsor PROME Web portals. Overall, this study found that there is a need for developing a trustworthy Web portal to systematically compile PROME for older patients to communicate well with their physicians.

### Acknowledgments

We would like to thank Dr Sunghee Tak for her valuable insights on patients' medication experience. We also thank the University of Tennessee Health Science Center, Seoul National University, and Seoul National University Research Institute of Pharmaceutical Sciences for their financial support.

### Conflicts of Interest

None declared.

### References

1. Manary MP, Boulding W, Staelin R, Glickman SW. The patient experience and health outcomes. N Engl J Med 2013 Jan 17;368(3):201-203. [doi: 10.1056/NEJMp1211775] [Medline: 23268647]
2. Giordano LA, Elliott MN, Goldstein E, Lehrman WG, Spencer PA. Development, implementation, and public reporting of the HCAHPS survey. Med Care Res Rev 2009 Jul 28;67(1):27-37. [doi: 10.1177/1077558709341065] [Medline: 19638641]
3. Morris B, Jahangir A, Sethi M. Patient satisfaction: an emerging health policy issue. Am Acad Orthop Surg 2013;6:7-9 [FREE Full text]
4. Doran T, Fullwood C, Gravelle H, Reeves D, Kontopantelis E, Hiroeh U, et al. Pay-for-performance programs in family practices in the United Kingdom. N Engl J Med 2006 Jul 27;355(4):375-384. [doi: 10.1056/NEJMsa055505] [Medline: 16870916]
5. Aikin KJ, Swasy JL, Braman AC. Patient and Physician Attitudes and Behaviors Associated With DTC Promotion of Prescription Drugs: Summary of FDA Survey Research Results. Executive Summary November 19, 2004. Silver Spring, MD: US Department of Health and Human Services, Food and Drug Administration, Center for Drug Evaluation and Research; 2004. URL: http://www.fda.gov/downloads/Drugs/ScienceResearch/ResearchAreas/DrugMarketingAdvertisingandCommunicationsResearch/UCM152890.pdf [accessed 2016-07-14] [WebCite Cache ID 6izKfRe2v]
6. Ask a Patient. Arlington, VA: Consumer Health Resource Group LLC; 2016. URL: http://www.askapatient.com/ [accessed 2016-07-12] [WebCite Cache ID 6ix2wbuDZ]
7. DrugRatingz. Sunnyvale, CA: Ratingz Inc; 2016. URL: http://www.drugratingz.com/ [accessed 2016-07-12] [WebCite Cache ID 6ix2zHVQ6]
8. WebMD. Atlanta, GA: WebMD, LLC; 2016. URL: http://www.webmd.com/ [accessed 2016-07-12] [WebCite Cache ID 6ix32R1nV]
9. ConsumerReports. Yonkers, NY: Consumer Reports; 2016. URL: http://www.consumerreports.org/cro/index.htm [accessed 2016-07-12] [WebCite Cache ID 6ix35Z19c]



XSL•FO
RenderX

10. 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;10(3):e15 [FREE Full text] [doi: 10.2196/jmir.1053] [Medline: 18504244]
11. Ru B, Harris K, Yao L. A content analysis of patient-reported medication outcomes on social media. 2015 Presented at: 2015 IEEE International Conference on Data Mining Workshop (ICDMW); November 14-17, 2015; Atlantic City, NJ. [doi: 10.1109/ICDMW.2015.150]
12. AlRuthia Y. The Value of Online Medication Rating Systems to Older Adults and Their Association with Self-Reported Outcomes [dissertation]. Memphis, TN: University of Tennessee Health Science Center; 2015. URL: http://etd.uthsc.edu/WORLD-ACCESS/AlRuthia/2015-002-AlRuthia.pdf [accessed 2016-03-11] [WebCite Cache ID 6fvB6hiXU]
13. Grover R, Vriens M. The Handbook of Marketing Research: Uses, Misuses, and Future Advances. Thousand Oaks, CA: Sage Publications; 2006.
14. Bardes CL. Defining "patient-centered medicine". N Engl J Med 2012 Mar 1;366(9):782-783. [doi: 10.1056/NEJMp1200070] [Medline: 22375968]
15. Levinson W, Lesser CS, Epstein RM. Developing physician communication skills for patient-centered care. Health Aff (Millwood) 2010 Jul;29(7):1310-1318 [FREE Full text] [doi: 10.1377/hlthaff.2009.0450] [Medline: 20606179]
16. Mavis B, Holmes RM, Jorgenson S, Coffey J, Anand N, Bulica E, et al. Patient participation in clinical encounters: a systematic review to identify self-report measures. Health Expect 2015 Dec;18(6):1827-1843. [doi: 10.1111/hex.12186] [Medline: 24611995]
17. Kiousis S. Public trust or mistrust? Perceptions of media credibility in the information age. Mass Commun Soc 2001 Nov;4(4):381-403. [doi: 10.1207/S15327825MCS0404_4]
18. Fritch JW, Cromwell RL. Evaluating Internet resources: identity, affiliation, and cognitive authority in a networked world. J Am Soc Inf Sci 2001;52(6):499-507. [doi: 10.1002/asi.1081]
19. Silver MP. Patient perspectives on online health information and communication with doctors: a qualitative study of patients 50 years old and over. J Med Internet Res 2015;17(1):e19 [FREE Full text] [doi: 10.2196/jmir.3588] [Medline: 25586865]
20. Beck F, Richard J, Nguyen-Thanh V, Montagni I, Parizot I, Renahy E. Use of the internet as a health information resource among French young adults: results from a nationally representative survey. J Med Internet Res 2014;16(5):e128 [FREE Full text] [doi: 10.2196/jmir.2934] [Medline: 24824164]
21. Zulman DM, Kirch M, Zheng K, An LC. Trust in the internet as a health resource among older adults: analysis of data from a nationally representative survey. J Med Internet Res 2011;13(1):e19 [FREE Full text] [doi: 10.2196/jmir.1552] [Medline: 21324832]
22. Zickuhr K, Madden M. Older adults and internet use. Washington, DC: Pew Internet & American Life Project; 2012 Jun 06. URL: http://www.sainetz.at/dokumente/Older_adults_and_internet_use_2012.pdf [accessed 2016-05-16] [WebCite Cache ID 6hXi86Xcg]
23. Johnson RW, Mommaerts C. Unemployment Rate Hits All-Time High for Adults Age 65 and Older. Washington, DC: Urban Institute; 2009 Mar 12. URL: http://www.taxpolicycenter.org/uploadedpdf/411846_recessionandolderworkersfactsheetmarch2009.pdf [accessed 2016-03-11] [WebCite Cache ID 6fvBFiGDe]
24. Joyce GF, Keeler EB, Shang B, Goldman DP. The lifetime burden of chronic disease among the elderly. Health Aff (Millwood) 2005 Sep;24 Suppl 2:W5R18-W5R29 [FREE Full text] [doi: 10.1377/hlthaff.w5.r18] [Medline: 16186148]
25. Metzger MJ, Flanagin AJ. Credibility and trust of information in online environments: the use of cognitive heuristics. J Pragmatics 2013 Dec;59:210-220. [doi: 10.1016/j.pragma.2013.07.012]
26. Westerwick A. Effects of sponsorship, web site design, and Google ranking on the credibility of online information. J Comput-Mediat Commun 2013 Jan 28;18(2):80-97. [doi: 10.1111/jcc4.12006]
27. Fox S, Rainie L. Vital decisions: how internet users decide what information to trust when they or their loved ones are sick. Washington, DC: Pew Internet & American Life Project; 2002 Mar 22. URL: http://www.pewinternet.org/files/old-media//Files/Reports/2002/PIP_Vital_Decisions_May2002.pdf.pdf [accessed 2016-03-11] [WebCite Cache ID 6fvBJVzBd]
28. Fogg BJ, Soohoo D, Danielson D, Marable L, Stanford J, Tauber ER. Fogg BJ, Soohoo C, Danielson D, Marable L, Stanford J, Tauber ER. 2002. How Do People Evaluate a Web Site's Credibility? URL: http://simson.net/ref/2002/stanfordPTL.pdf [accessed 2016-03-11] [WebCite Cache ID 6fvBMmtmg]
29. Freeman K, Spyridakis J. An examination of factors that affect the credibility of online health information. Tech Commun 2004;51(2):239-263 [FREE Full text]
30. Masino C, Lam TC. Choice of rating scale labels: implication for minimizing patient satisfaction response ceiling effect in telemedicine surveys. Telemed J E Health 2014 Dec;20(12):1150-1155. [doi: 10.1089/tmj.2013.0350] [Medline: 25379838]
31. Liana E. SurveyMonkey Blog. Palo Alto, CA: SurveyMonkey Inc; 2012 Jan 24. Words Speak Louder Than Numbers URL: https://www.surveymonkey.com/blog/2012/01/24/words-speak-louder-than-numbers/ [accessed 2016-05-19] [WebCite Cache ID 6hcE3M3rn]
32. Menold N, Kaczmirek L, Lenzner T, Neusar A. How do respondents attend to verbal labels in rating scales? Field Methods 2013 Oct 23;26(1):21-39. [doi: 10.1177/1525822X13508270]




XSL•FO
RenderX

## Abbreviations

**DTCA:** direct-to-consumer advertising
**FDA:** US Food and Drug Administration
**PROME:** patients' reports of medication experience

*Edited by G Eysenbach; submitted 29.03.16; peer-reviewed by I Montagni, M Rigby, JW Fritch; comments to author 12.05.16; revised version received 24.05.16; accepted 05.07.16; published 22.07.16.*

*Please cite as:*
*Hong SH, Lee W, AlRuthia Y*
*Health Care Applicability of a Patient-Centric Web Portal for Patients' Medication Experience*
*J Med Internet Res 2016;18(7):e202*
*URL: http://www.jmir.org/2016/7/e202/*
*doi:10.2196/jmir.5813*
*PMID:27450362*

©Song Hee Hong, Woojung Lee, Yazed AlRuthia. Originally published in the Journal of Medical Internet Research (http://www.jmir.org), 22.07.2016. This is an open-access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/2.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on http://www.jmir.org/, as well as this copyright and license information must be included.



XSL•FO
RenderX

Original Paper

# Online Health Information-Seeking Behavior and Confidence in Filling Out Online Forms Among Latinos: A Cross-Sectional Analysis of the California Health Interview Survey, 2011-2012

Mariaelena Gonzalez[1,2*], PhD; Ashley Sanders-Jackson[3*], PhD; Jason Emory[4], PhD

[1]Public Health, University of California, Merced, Merced, CA, United States

[2]Health Sciences Research Institute, University of California, Merced, Merced, CA, United States

[3]Department of Advertising and PR, School of Communication Arts and Sciences, Michigan State University, East Lansing, MI, United States

[4]Department of Psychology & Child Development, California State University, Stanislaus, Turlock, CA, United States

[*]these authors contributed equally

**Corresponding Author:**
Mariaelena Gonzalez, PhD
Public Health
University of California, Merced
5200 N Lake Road
Merced, CA, 95343
United States
Phone: 1 650 283 8307
Fax: 1 209 228 4007
Email: mgonzalez82@ucmerced.edu

## Abstract

**Background:** Health information is increasingly being disseminated online, but there is a knowledge gap between Latinos and non-Hispanic whites, particularly those whose English language proficiency is poor, in terms both of online health information-seeking behavior and computer literacy skills. This knowledge gap may also exist between US- and foreign-born Latinos.

**Objective:** The specific aim of this study was to examine Internet use, online health information-seeking behavior, and confidence in filling out online forms among Latinos, particularly as it relates to health-risk behaviors. We then stratified our sample by nativity.

**Methods:** We used the adult population file of the 2011-2012 California Health Interview Survey, analyzing Internet use, online health information-seeking behavior, and confidence in filling out online forms using binary logistic regression among Latinos and whites (N=27,289), Latinos (n=9506), and Latinos who use the Internet (n=6037).

**Results:** Foreign-born Latinos (OR 0.71, 95% CI 0.58-0.88, *P*=.002) have lower odds of engaging in online health information-seeking behavior, and higher odds (OR 2.90, 95% CI 2.07-4.06, *P*<.001) of reporting a lack of confidence in filling out online forms compared to US-born Latinos. Correlates of online health information-seeking behavior and form confidence varied by nativity.

**Conclusions:** Latinos, particularly foreign-born individuals, are at an increased risk of being left behind as the move to increase online content delivery and care expands. As online health information dissemination and online health portals become more popular, the impact of these sites on Latino gaps in coverage and care should be considered.

***(J Med Internet Res 2016;18(7):e184)*** doi:10.2196/jmir.5065

**KEYWORDS**

communications media; social media; health information-seeking behavior; online forms; Hispanic Americans



XSL•FO
RenderX

## Introduction

Health information and marketing is increasingly being disseminated and collected via the Internet [1]; therefore, understanding who is and is not engaging in online health information-seeking behavior is crucial to improving health. Health information-seeking behavior is the third most popular use of the Internet among adults [2,3]. The use of online and mobile interventions shows promising results as a way to implement large-scale behavioral changes [4]. As a result, tracking online health information-seeking behavior is particularly important because online health information-seeking behavior is associated with risk behaviors in the general population [5]. However, this association between online health information-seeking behavior and Internet use varies by behavior. Decreased fruit and vegetable consumption are associated with decreased use of the Internet for health information-seeking behavior, whereas others, such as smoking, are associated with increased online health information-seeking behavior in the general US population [5]. Additionally, it is important to understand competencies in performance of online tasks—such as the ability to fill out online forms—because insurers, hospitals, and other government programs are using online forms and online apps as ways of gathering information and communicating with their clients [6].

There are significant geographic, social, economic, and racial/ethnic disparities in online health information-seeking behavior [7,8]. In particular, although more than 75% of Latinos use the Internet or email at least occasionally [9], Latinos are less likely than other groups to go online for health information and to use online health services such as Internet portals [3,10-12]. Further, differences in English language literacy, eHealth literacy, and background knowledge may also affect how much benefit people receive from health-related information online [13-15]. We define eHealth literacy as "the ability to seek, find, understand, and appraise health information from electronic sources and apply the knowledge gained to addressing or solving a health problem" [16]. Although computer literacy is part of eHealth literacy, it is a fairly complex construct that allows scholars to understand how existing disparities can be exacerbated be electronic tools due to the ability of the existing system to reinforce structural inequalities [14].

Although English language proficiency has been increasing among Latinos overall [17], there continues to be disparities in eHealth literacy and health-related background knowledge. For example, Latinos are less likely than other groups to have high computer use efficacy and they have a lower activation rate for online health systems than whites [18,19]. In addition, in a study of parents in Florida, Latino parents had less success accessing an eHealth portal for help with their special needs children [4]. Latinos with poorer English language proficiency are less likely to seek health information online [15] and older Latinos with poorer English tend to have lower eHealth literacy [5]. Further, Latino women are less likely to seek health information online than white women [10], which may lead to online health information knowledge gaps. For example, Latinos typically have lower levels of cancer-related health information than whites do [20-22].

When examining Latinos, it is important to consider that there exists significant differences between US- and foreign-born Hispanics in the United States. Latinos born in the United States are generally younger and have higher socioeconomic status (as measured by education and income levels) than non-US-born Latinos [23,24]. With each successive generation born in the United States, English language dominance increases as Spanish usage declines. English language proficiency and nativity are also significant when considering Hispanic Internet usage in general [17,23]. A Pew survey shows that half of all American Hispanic Internet users are US-born and the majority of American Hispanic Internet users are either English dominant or bilingual [9].

Although some recent studies have been done on online health information-seeking behavior among Latinos to determine the profile of Latino online health information-seeking behavior, the majority of these studies have been small or limited to particular localities (ie, boroughs of New York or Puerto Rico) [11,25-28]. One study examined a national sample of Latinos [29]. We found no studies on use of the Internet for health information-seeking behavior by Latinos who engage in health-risk behavior. Furthermore, we found no study that stratifies Latinos by nativity in order to understand how the correlates of online health information-seeking behavior differs between foreign- and US-born Latinos. Understanding how online health information-seeking behavior by Latinos varies by nativity is important because the use of media for health information and prevalent modes of media use among Latinos differs according to demographic characteristics [29,30].

California is the most populous US state; in 2015, 38% of its population was Hispanic or Latino as compared to 17.4% in the US population overall [31]. Due to the high percentage of Hispanics in California, research related to Hispanic or Latino individuals often sample from there specifically [32-34]. As a result, understanding Internet access and online health information-seeking behavior of Latinos in California will help to predict national trends. When studying Latinos in California, it is particularly important to examine US- versus foreign-born Latinos because there is a significant demographic divide between Latinos in the United States and in California specifically; in 2011, the average age of US-born Latinos in California was 18 years and 72% of this population were younger than age 30 [35]. This mirrors the demographic shift that the Hispanic population in the United States has been undergoing over the last three decades. As a result, although Hispanics are still perceived as an immigrant population, in 2013, 64.8% of all US Hispanics were born in the United States [36].

The purpose of this study is to examine Internet use, online health information-seeking behavior, and confidence in filling out online forms (form confidence) among Latinos in California, particularly as it relates to health-risk behaviors. This study (1) examines whether or not disparities between Internet use, online health information-seeking behavior, and lack of confidence in filling out online forms varies between non-Hispanic whites



XSL•FO
RenderX

and Latinos, (2) then analyzes Internet use and online health information-seeking behavior and lack of confidence in filling out online forms among Latinos, and (3) examines the correlates of health information-seeking behavior and confidence in filling out online forms among US- and foreign-born Latinos, stratifying by nativity.

## Methods

Data came from the adult population file of the 2011-2012 California Health Interview Survey (CHIS), which is a representative sample of noninstitutionalized California state adult population aged 18 and older [37]. We analyzed Latinos and non-Hispanic whites (N=27,289), all Latinos (n=9506), and the subsample of Latinos who reported using the Internet in the last 12 months (n=6037). Additionally we stratified our analyses of Latinos by nativity.

### Primary Outcomes

Our three primary outcomes of interest were (1) ever used the Internet, defined as having ever used the Internet including email and social media sites; (2) health information-seeking behavior in the last 12 months, defined as seeking health or medical information online, including information about disease symptoms, diet or nutrition, physical activity, health care providers, and health insurance plans; and (3) lack of confidence in filling out an online application on their own (not at all confident or not too confident=1, somewhat confident or very confident=0).

### Sociodemographic and Health Risk Variables

We controlled for age (18-24, 25-34, 35-44, 45-64, and ≥65 years), gender (female=1, male=0), education (less than high school, high school, some college, and college or greater), poverty level (poor: 0%-99% of the federal poverty level [FPL], near poor 100%-199% of the FPL, above poor: ≥200% of the FPL), marital status (married=1), employment status (employment=1), and living in a household with kids (children=1). We collapsed the first three categories of self-rated health (excellent, very good, good) into a single category (good) and compared that to all other responses. Additionally, we controlled for having a usual source of care other than the emergency department, being currently insured, and having been diagnosed with one of eight chronic diseases (asthma, diabetes or prediabetes, heart disease, stroke, arthritis, gout, or lupus).

Obesity was defined as having a body mass index (BMI) in the obese range according to Centers for Disease Control and Prevention (CDC) definitions. “Smoker” was defined as someone who has smoked at least 100 cigarettes in their lifetime and reported currently smoking some days or every day. Frequent binge drinking was defined as binge drinking (≥5 drinks for men and ≥4 drinks for women) monthly or more often. The majority of respondents reported eating less than seven servings of vegetables per week (respondents were asked how often they eat vegetables per week), so we dichotomized vegetable consumption into those who consumed vegetables six or fewer times per week and those who consumed vegetables seven or more times per week. We dichotomized variables representing unfavorable eating habits into consuming two or more servings of soda per week and consuming two or more servings of fast food per week. Additionally we dichotomized the sample into individuals who walked less than 150 minutes per week (for work or exercise).

### Other Measures

In models comparing Latinos to whites, we controlled for being Latino (Latino=yes, non-Hispanic white=no). As previously discussed, there exists significant demographic and cultural differences between US- and foreign-born Hispanics in the United States [23,24]. Therefore, we used the US Census definition of nativity, dividing Latino respondents into US- and foreign-born individuals. Because language use varies between US- and foreign-born Latinos and is related to Internet use among Latinos in the United States [9], we controlled for acculturation using English language proficiency and language in which media is consumed. We measured a respondent’s lack of acculturation via the proxies (1) respondent’s English proficiency is low and (2) respondent only consumes media in a non-English language.

We controlled for neighborhood-level factors with two social capital measures, trust/safety and civic engagement, constructed from individual-level social capital variables in the CHIS. Trust/safety measures respondents’ perceptions that people in their neighborhood can be trusted and are willing to help one another, their feelings of safety in their neighborhood, and perception that neighbors watch out for the safety of children in the neighborhood. Civic engagement measures volunteering in the community and attending meetings dealing with community problems during the past year. These measures were created by averaging the individual items in the CHIS following a factor analysis to ensure that all items loaded on a single factor.

### Analysis

We first conducted three logistic regressions comparing Latinos to whites to test if Latinos odds of (1) using the Internet, (2) engaging in online health information-seeking behavior, and (3) lack of confidence in filling out online forms were lower than non-Hispanic whites. Our sample population for this analysis were non-Hispanic whites and Latinos. We then used logistic regressions to (1) test whether US-born Latinos were more likely to use the Internet compared foreign-born Latinos, (2) examine the correlates of online health information-seeking behavior, and (3) examine the characteristics of Latino Internet users who were not confident filling out online forms. After we conducted the last two Latino analyses, we stratified our sample by nativity. All analyses were weighted according to CHIS directions.

## Results

In all, 6037 of 9506 (69.95%) Latinos reported using the Internet in the last year (Table 1). Of these 6037 Latinos, 3190 (53.36%) reported engaging in online health information-seeking behavior and 998 of 6037 (15.64%) reported a lack of confidence in filling out online forms.



XSL•FO
RenderX

**Table 1.** Descriptive characteristics of all Latinos and Latinos who use the Internet in the California Health Interview Survey (2011-2012).

| Characteristics | Latinos (n=9506) | Latinos who use the Internet (n=6037) |
|---|---|---|
| Use the Internet, n (%)[a] | 6037 (69.95) | — |
| Engage in online health information-seeking behavior, n (%) | — | 3190 (53.36) |
| Lack online form confidence (not at all confident or not too confident filling out online forms), n (%) | — | 998 (15.64) |
| **Nativity, n (%)** | | |
| Foreign-born | 5425 (54.93) | 2672 (42.33) |
| US-born | 4081 (45.07) | 3365 (57.67) |
| **Acculturation, n (%)** | | |
| English proficiency is low | 3659 (37.14) | 1225 (20.52) |
| Only consumes media in the non-English language | 2482 (24.77) | 720 (11.62) |
| **Age (years), n (%)** | | |
| 18-24 | 1332 (18.70) | 1296 (26.17) |
| 25-34 | 1474 (22.25) | 1200 (26.16) |
| 35-44 | 1923 (22.58) | 1338 (22.66) |
| 45-64 | 3354 (28.77) | 1819 (22.30) |
| ≥65 | 1423 (7.70) | 384 (2.72) |
| Female, n (%) | 5531 (50.53) | 3376 (49.97) |
| **Education, n (%)** | | |
| Below high school | 3267 (34.98) | 965 (18.71) |
| High school | 2676 (28.71) | 2020 (33.92) |
| Some college | 2125 (22.13) | 1753 (28.32) |
| Bachelors or higher | 1438 (14.17) | 1299 (19.05) |
| **Poverty level, n (%)** | | |
| Poor (0%-99% of the FPL) | 2939 (29.09) | 1437 (23.05) |
| Near poor (100%-199% of the FPL) | 2768 (29.15) | 1567 (26.87) |
| Above poor (≥200% of the FPL) | 3799 (41.76) | 3033 (50.09) |
| Married, n (%) | 4638 (46.92) | 2802 (42.14) |
| Employed, n (%) | 5890 (58.90) | 3641 (64.10) |
| Lives in a family with children, n (%) | 3881 (44.00) | 2645 (44.03) |
| Self-rated health is good to excellent, n (%) | 6540 (72.49) | 4845 (81.46) |
| Has usual sources of care, n (%) | 7553 (75.74) | 4892 (77.11) |
| Currently insured, n (%) | 7022 (70.05) | 4485 (71.26) |
| Chronic diseases, n(%) | 3276 (26.73) | 1579 (20.18) |
| **Risk behavior, n (%)** | | |
| Obese | 3240 (32.55) | 1810 (28.73) |
| Smoker | 1014 (12.27) | 661 (12.30) |
| Binge drinks once a month or more often | 886 (12.51) | 742 (15.39) |
| Drinks soda ≥2 times per week | 3247 (39.55) | 2185 (41.16) |
| Eats ≥2 servings of fast food per week | 3600 (44.34) | 2708 (50.93) |
| Eats <7 servings of vegetables per week | 6012 (65.79) | 3708 (64.54) |
| Walks <150 minutes per week | 8050 (84.14) | 5146 (84.11) |



XSL•FO
RenderX

| Characteristics | Latinos (n=9506) | Latinos who use the Internet (n=6037) |
|---|---|---|
| **Social capital, mean (SD)** | | |
| Trust and safety (continuous) | 2.94 (0.01) | 2.94 (0.01) |
| Civic engagement (continuous) | 0.17 (0.003) | 0.21 (0.004) |

[a] Percentages are weighted according to California Health Interview Survey directions in order to provide California population estimates.

## White Versus Latino Knowledge Gap

We found that Latinos were less likely to have ever used the Internet (OR 0.48, 95% CI 0.41-0.56, *P*<.001), to have engaged in online health information-seeking behavior (OR 0.71, 95% CI 0.63-0.80, *P*<.001), and were more likely to say that they were not at all or not too confident they could fill out an online application on their own (OR 1.7, 95% CI 1.36-2.12, *P*<.001) compared to non-Hispanic whites.

## Latino Internet Use

We found no significant differences in Internet use between US- and foreign-born Latinos when examining all Latinos (both US- and foreign-born, Table 2). Individuals with low English proficiency (OR 0.33, 95% CI 0.23-0.47, *P*<.001) or who consumed media in a non-English language (OR 0.46, 95% CI 0.35-0.60, *P*<.001) had lower odds of using the Internet. Individuals ages 25-34 (OR=0.19, 95% CI 0.12-0.31, *P*<.001) had lower odds than those aged 18 to 24 years; Latinos aged 65 or older (OR 0.01, 95% CI 0.01-0.02, *P*<.001) had significantly lower odds of using the Internet. Lower education and poverty levels were associated with never having used the Internet. Women (OR 1.28, 95% CI 1.04-1.59, *P*=.02), employed individuals (OR 1.29, 95% CI 1.04-1.60, *P*=.02), individuals living in a family with children (OR 1.37, 95% CI 1.07-1.75, *P*=.01), and with good self-rated health (OR 1.43, 95% CI 1.10-1.85, *P*=.008) had higher odds of ever using the Internet, whereas individuals with a chronic disease (OR 0.69, 95% CI 0.57-0.84, *P*<.001) had lower odds of ever using the Internet. When examining health-risk behaviors, individuals who were obese (OR 0.79, CI 0.64-0.98, *P*=.03) had lower odds of ever having used the Internet, whereas Latinos who were frequent binge drinkers (OR 1.49, 95% CI 1.07-2.08, *P*=.02), or who ate two or more servings of fast food (OR 1.65, 95% CI 1.31-2.07, *P*<.001) had higher odds of using the Internet. When examining social integration, individuals with higher levels of trust had lower odds of ever having using the Internet (OR 0.81, 95% CI 0.66-0.99, *P*=.04), whereas individuals who were civically engaged had higher odds (OR 3.13, 95% CI 2.18-4.50, *P*=.001) of ever having used the Internet.

## Online Health Information-Seeking Behavior

When examining online health information-seeking behavior among all Latinos (Table 3), we found that foreign-born Latinos (OR 0.71, 95% CI 0.58-0.88, *P*=.002) had lower odds of engaging in online health information-seeking behavior than US-born Latinos. Individuals aged 65 years or older (OR 0.54, 95% CI 0.40-0.75, *P*<.001), who had lower levels of education (below high school: OR 0.33, 95% CI 0.24-0.46, *P*<.001; high school: OR 0.53, 95% CI 0.41-0.67, *P*<.001; some college: OR 0.72, 95% CI 0.59-0.87, *P*=.001), and who were poor (OR 0.75, 95% CI 0.61-0.93, *P*=.01) had lower odds of engaging in online health information-seeking behavior. Women (OR 1.50, 95% CI 1.27-1.77, *P*<.001), individuals with a usual source of care (OR 1.43, 95% CI 1.14-1.79, *P*=.01), and individuals with a chronic disease (OR 1.26, 95% CI 1.02-1.57, *P*=.03) had higher odds of online health information-seeking behavior. When examining health-risk behaviors, individuals who smoked (OR 0.78, 95% CI 0.62-1.00, *P*=.047) had lower odds of health information-seeking behavior, whereas individuals who frequently binge drank had higher odds (OR 1.40, 95% CI 1.11-1.78, *P*=.006) of online health information-seeking behavior.

When stratifying by nativity (Table 4) we found that US-born Latinos who had a usual source of care (OR 1.56, 95% CI 1.17-2.08, *P*=.003) and binge drinking (OR 1.53, 95% CI 1.14-2.06, *P*=.006) had higher odds of online health information-seeking behavior, whereas these items were not associated with online health information-seeking behavior among foreign-born Latinos. Smokers had lower odds (OR 0.65, 95% CI 0.43-0.98, *P*=.03) of online health information-seeking behavior among foreign-born Latinos, but this was not associated with online health information-seeking behavior among US-born individuals.



XSL•FO
RenderX

**Table 2.** Logistic regression: correlates of Internet use among Latinos in California from CHIS 2011-2012 (N=9474).

| Characteristics | OR (95% CI) | *P* |
|---|---|---|
| **Nativity** | | |
| Foreign-born | 0.82 (0.60-1.12) | .21 |
| **Acculturation** | | |
| English proficiency is low | 0.33 (0.23-0.47) | <.001 |
| Only consumes media in non-English language | 0.46 (0.35-0.60) | <.001 |
| **Age (years)** | | |
| 25-34 | 0.19 (0.12-0.31) | <.001 |
| 35-44 | 0.13 (0.08-0.22) | <.001 |
| 45-64 | 0.06 (0.04-0.10) | .001 |
| ≥65 | 0.01 (0.01-0.02) | <.001 |
| Female | 1.28 (1.04-1.59) | .02 |
| **Education** | | |
| Below high school | 0.12 (0.08-0.20) | <.001 |
| High school | 0.36 (0.23-0.56) | <.001 |
| Some college | 0.51 (0.32-0.82) | .006 |
| **Poverty level** | | |
| Poor (0%-99% of the FPL) | 0.46 (0.35-0.61) | <.001 |
| Near poor (100%-199% of the FPL) | 0.73 (0.57-0.95) | .02 |
| Married | 1.03 (0.81-1.32) | .81 |
| Employed | 1.29 (1.04-1.60) | .02 |
| Lives in a family with children | 1.37 (1.07-1.75) | .01 |
| Self-rated health is good to excellent | 1.43 (1.10-1.85) | .008 |
| Has usual sources of care | 1.20 (0.95-1.53) | .13 |
| Currently insured | 1.05 (0.81-1.37) | .68 |
| Chronic diseases | 0.69 (0.57-0.84) | <.001 |
| **Risk behavior** | | |
| Obese | 0.79 (0.64-0.98) | .03 |
| Smoker | 0.82 (0.61-1.10) | .18 |
| Binge drinks once a month or more often | 1.49 (1.07-2.08) | .02 |
| Drinks soda ≥2 times per week | 1.09 (0.84-1.42) | .50 |
| Eats ≥2 servings of fast food per week | 1.65 (1.31-2.07) | <.001 |
| Eats <7 servings of vegetables per week | 0.92 (0.71-1.18) | .49 |
| Walks <150 minutes per week | 1.04 (0.79-1.36) | .79 |
| **Social capital** | | |
| Trust and safety | 0.81 (0.66-0.99) | .04 |
| Civic engagement | 3.13 (2.18-4.50) | <.001 |
| Constant | 204.57 (86.81-482.07) | <.001 |



XSL•FO
RenderX

**Table 3.** Logistic regression: correlates of online health information-seeking behavior and lack of form confidence among Latinos in California from CHIS 2011-2012 (n=6035).

| Characteristics | Online health information-seeking behavior | | Lack of form confidence | |
|---|---|---|---|---|
| | OR (95% CI) | *P* | OR (95% CI) | *P* |
| **Nativity** | | | | |
| Foreign-born | 0.71 (0.58-0.88) | .002 | 2.90 (2.07-4.06) | <.001 |
| **Acculturation** | | | | |
| English proficiency is low | 0.81 (0.61-1.07) | .14 | 2.51 (1.89-3.34) | <.001 |
| Only consumes media in non-English language | 0.74 (0.54-1.03) | .08 | 1.46 (1.05-2.03) | .03 |
| **Age (years)** | | | | |
| 25-34 | 1.24 (0.94-1.62) | .12 | 0.87 (0.57-1.34) | .53 |
| 35-44 | 1.04 (0.76-1.40) | .82 | 1.25 (0.79-1.99) | .33 |
| 45-64 | 0.86 (0.66-1.13) | .28 | 1.10 (0.71-1.71) | .67 |
| ≥65 | 0.54 (0.40-0.75) | <.001 | 2.03 (0.94-4.38) | .07 |
| Female | 1.50 (1.27-1.77) | <.001 | 1.17 (0.90-1.51) | .23 |
| **Education** | | | | |
| Below high school | 0.33 (0.24-0.46) | <.001 | 2.67 (1.76-4.05) | <.001 |
| High school | 0.53 (0.41-0.67) | <.001 | 1.89 (1.26-2.84) | .003 |
| Some college | 0.72 (0.59-0.87) | .001 | 1.11 (0.71-1.73) | .63 |
| **Poverty level** | | | | |
| Poor (0%-99% of the FPL) | 0.75 (0.61-0.93) | .01 | 1.47 (0.98-2.22) | .06 |
| Near poor (100%-199% of the FPL) | 0.91 (0.74-1.11) | .34 | 1.77 (1.28-2.45) | .001 |
| Married | 1.01 (0.82-1.25) | .92 | 1.23 (0.92-1.65) | .16 |
| Employed | 0.99 (0.83-1.18) | .92 | 0.96 (0.74-1.25) | .78 |
| Lives in a family with children | 0.95 (0.77-1.18) | .66 | 0.91 (0.67-1.25) | .56 |
| Self-rated health is good to excellent | 1.03 (0.82-1.29) | .81 | 0.62 (0.44-0.88) | .01 |
| Has usual sources of care | 1.43 (1.14-1.79) | .002 | 0.74 (0.52-1.04) | .08 |
| Currently insured | 1.06 (0.86-1.32) | .58 | 0.88 (0.61-1.26) | .47 |
| Chronic diseases | 1.26 (1.02-1.57) | .03 | 0.85 (0.60-1.21) | .37 |
| **Risk behavior** | | | | |
| Obese | 0.93 (0.76-1.12) | .44 | 0.99 (0.75-1.30) | .92 |
| Smoker | 0.78 (0.62-1.00) | .047 | 1.15 (0.72-1.82) | .55 |
| Binge drinks once a month or more often | 1.40 (1.11-1.78) | .006 | 0.78 (0.52-1.17) | .23 |
| Drinks soda ≥2 times per week | 0.89 (0.74-1.07) | .21 | 1.12 (0.86-1.47) | .39 |
| Eats ≥2 servings of fast food per week | 1.08 (0.91-1.29) | .39 | 0.72 (0.54-0.95) | .02 |
| Eats <7 servings of vegetables per week | 0.88 (0.75-1.03) | .10 | 1.21 (0.93-1.57) | .15 |
| Walks <150 minutes per week | 1.08 (0.86-1.37) | .49 | 1.02 (0.74-1.41) | .91 |
| **Social capital** | | | | |
| Trust and safety | 0.89 (0.77-1.04) | .13 | 0.70 (0.58-0.83) | <.001 |
| Civic engagement | 2.64 (2.01-3.49) | <.001 | 0.69 (0.47-1.02) | .07 |
| Constant | 1.74 (0.93-3.23) | .08 | 0.15 (0.06-0.37) | <.001 |



XSL•FO
RenderX

**Table 4.** Logistic regression: correlates of online health information-seeking behavior among Latinos in California stratified by nativity (source: CHIS 2011-2012).

| Characteristics | US-born Latinos (n=3363) | | Foreign-born Latinos (n=2672) | |
|---|---|---|---|---|
| | OR (95% CI) | *P* | OR (95% CI) | *P* |
| **Acculturation** | | | | |
| English proficiency is low | 0.42 (0.17-1.00) | .049 | 0.78 (0.57-1.07) | .12 |
| Only consumes media in non-English language | 0.70 (0.29-1.67) | .42 | 0.74 (0.52-1.05) | .09 |
| **Age (years)** | | | | |
| 25-34 | 1.08 (0.77-1.50) | .66 | 1.40 (0.89-2.20) | .14 |
| 35-44 | 0.87 (0.54-1.40) | .55 | 1.21 (0.78-1.88) | .39 |
| 45-64 | 0.80 (0.57-1.14) | .21 | 0.91 (0.58-1.44) | .69 |
| ≥65 | 0.57 (0.37-0.89) | .01 | 0.47 (0.24-0.91) | .03 |
| Female | 1.41 (1.12-1.78) | .004 | 1.53 (1.17-1.99) | .002 |
| **Education** | | | | |
| Below high school | 0.33 (0.20-0.53) | <.001 | 0.36 (0.24-0.55) | <.001 |
| High school | 0.46 (0.33-0.64) | <.001 | 0.64 (0.45-0.91) | .01 |
| Some college | 0.66 (0.49-0.88) | .006 | 0.81 (0.57-1.14) | .22 |
| **Poverty level** | | | | |
| Poor (0%-99% of the FPL) | 0.77 (0.56-1.05) | .10 | 0.76 (0.53-1.09) | .13 |
| Near poor (100%-199% of the FPL) | 0.94 (0.69-1.29) | .72 | 0.94 (0.68-1.31) | .73 |
| Married | 1.16 (0.89-1.52) | .26 | 0.90 (0.67-1.21) | .50 |
| Employed | 1.06 (0.86-1.31) | .58 | 0.87 (0.67-1.12) | .28 |
| Lives in a family with children | 1.01 (0.75-1.36) | .94 | 0.88 (0.67-1.16) | .35 |
| Self-rated health is good to excellent | 1.15 (0.82-1.62) | .40 | 0.92 (0.66-1.27) | .60 |
| Has usual sources of care | 1.56 (1.17-2.08) | .003 | 1.27 (0.92-1.76) | .15 |
| Currently insured | 1.09 (0.81-1.47) | .55 | 1.07 (0.79-1.46) | .67 |
| Chronic diseases | 1.28 (0.99-1.67) | .06 | 1.21 (0.88-1.65) | .24 |
| **Risk behavior** | | | | |
| Obese | 0.87 (0.67-1.13) | .29 | 1.02 (0.77-1.35) | .89 |
| Smoker | 0.92 (0.67-1.25) | .58 | 0.65 (0.43-0.98) | .04 |
| Binge drinks once a month or more often | 1.53 (1.14-2.06) | .01 | 1.16 (0.77-1.75) | .47 |
| Drinks soda ≥2 times per week | 0.81 (0.64-1.02) | .07 | 1.01 (0.77-1.31) | .95 |
| Eats ≥2 servings of fast food per week | 1.06 (0.83-1.35) | .63 | 1.10 (0.86-1.42) | .44 |
| Eats <7 servings of vegetables per week | 0.94 (0.78-1.15) | .56 | 0.81 (0.62-1.05) | .10 |
| Walks <150 minutes per week | 1.13 (0.83-1.54) | .44 | 1.05 (0.74-1.49) | .79 |
| **Social capital** | | | | |
| Trust and safety | 0.85 (0.69-1.06) | .15 | 0.93 (0.75-1.14) | .47 |
| Civic engagement | 3.55 (2.34-5.38) | <.001 | 1.87 (1.20-2.91) | .01 |
| Constant | 1.61 (0.71-3.61) | .25 | 1.45 (0.55-3.79) | .44 |



XSL•FO
RenderX

**Table 5.** Logistic regression: correlates of lack of online form confidence (not at all confident or not too confident filling out online forms) among Latinos in California stratified by nativity (source: CHIS 2011-2012).

| Characteristics | US-born Latinos (n=3363) | | Foreign-born Latinos (n=2672) | |
|---|---|---|---|---|
| | OR (95% CI) | *P* | OR (95% CI) | *P* |
| **Acculturation** | | | | |
| English proficiency is low | 1.99 (0.74-5.3) | .17 | 2.56 (1.89-3.47) | <.001 |
| Only consumes media in the non-English language | 1.82 (0.4-8.21) | .43 | 1.49 (1.10-2.02) | .01 |
| **Age (years)** | | | | |
| 25-34 | 0.57 (0.28-1.19) | .13 | 1.54 (0.85-2.79) | .16 |
| 35-44 | 0.57 (0.22-1.45) | .23 | 2.14 (1.15-3.97) | .02 |
| 45-64 | 0.89 (0.46-1.74) | .74 | 1.68 (0.95-2.97) | .07 |
| ≥65 | 3.54 (1.40-8.96) | .008 | 1.39 (0.52-3.69) | .50 |
| Female | 1.30 (0.83-2.03) | .25 | 1.11 (0.79-1.56) | .53 |
| **Education** | | | | |
| Below high school | 9.35 (2.39-36.59) | .002 | 2.06 (1.34-3.16) | .001 |
| High school | 5.17 (1.42-18.80) | .01 | 1.51 (0.95-2.42) | .08 |
| Some college | 3.94 (1.10-14.11) | .04 | 0.81 (0.47-1.37) | .42 |
| **Poverty level** | | | | |
| Poor (0%-99% of the FPL) | 1.20 (0.64-2.23) | .57 | 1.57 (0.96-2.55) | .07 |
| Near poor (100%-199% of the FPL) | 1.55 (0.91-2.66) | .11 | 1.73 (1.17-2.56) | .006 |
| Married | 1.09 (0.55-2.16) | .80 | 1.27 (0.93-1.74) | .16 |
| Employed | 0.68 (0.45-1.04) | .08 | 1.10 (0.78-1.54) | .58 |
| Lives in a family with children | 1.10 (0.51-2.35) | .81 | 0.81 (0.58-1.14) | .22 |
| Self-rated health is good to excellent | 0.41 (0.21-0.81) | .01 | 0.72 (0.50-1.04) | .08 |
| Has usual sources of care | 0.49 (0.24-1.00) | .049 | 0.87 (0.59-1.26) | .45 |
| Currently insured | 0.95 (0.49-1.84) | .88 | 0.84 (0.59-1.21) | .36 |
| Chronic diseases | 1.07 (0.57-2.02) | .83 | 0.82 (0.55-1.23) | .34 |
| **Risk behavior** | | | | |
| Obese | 0.80 (0.50-1.28) | .34 | 1.08 (0.77-1.50) | .66 |
| Smoker | 1.44 (0.65-3.20) | .36 | 1.03 (0.64-1.65) | .90 |
| Binge drinks once a month or more often | 0.60 (0.29-1.23) | .16 | 0.92 (0.56-1.51) | .75 |
| Drinks soda ≥2 times per week | 0.88 (0.58-1.32) | .53 | 1.18 (0.87-1.60) | .28 |
| Eats ≥2 servings of fast food per week | 1.06 (0.71-1.58) | .78 | 0.62 (0.44-0.87) | .01 |
| Eats <7 servings of vegetables per week | 1.25 (0.73-2.15) | .40 | 1.19 (0.90-1.57) | .22 |
| Walks less than<150 minutes per week | 1.18 (0.65-2.16) | .58 | 0.99 (0.67-1.45) | .95 |
| **Social capital** | | | | |
| Trust and safety | 0.78 (0.54-1.14) | .20 | 0.69 (0.55-0.87) | .002 |
| Civic engagement | 0.44 (0.20-0.97) | .04 | 0.82 (0.52-1.29) | .39 |
| Constant | 0.08 (0.01-0.58) | .01 | 0.29 (0.08-1.03) | .06 |



XSL•FO
RenderX

### Lack of Confidence in Filling Out Online Forms

Lack of confidence in filling out online forms was higher among Latinos with low English proficiency (Table 3, OR 2.51, 95% CI 1.89-3.34, *P*<.001) or those who only consumed media in a foreign language (OR 1.46, 95% CI 1.05-2.03, *P*=.03). Individuals with a high school or lower education level (below high school: OR 2.67, 95% CI 1.76-4.05, *P*<.001; high school: OR 1.89, 95% CI 1.26-2.84, *P*=.003) or who were near poor (OR 1.77, 1.28-2.45, *P*=.001) had higher odds of lacking confidence. Individuals with good self-rated health (OR 0.62, 95% CI 0.44-0.88, *P*=.01), who frequently ate fast food (OR 0.72, 95% CI 0.54-0.95, *P*=.03), or with higher levels of trust and safety were less likely to express a lack of confidence in filling out online forms.

When we stratified our sample by nativity (Table 5), we found that among US-born Latinos, elderly (≥65 years) individuals had higher odds (OR 3.54, 95% CI 1.40-8.96, *P*=.008) of expressing a lack of confidence. When examining socioeconomic status, we found that only education was related to lack of confidence. Although individuals with less than college had higher odds of expressing a lack of confidence filling out forms, Latinos with less than a high school education had more than 800% higher odds (OR 9.53, 95% CI 2.39-36.59, *P*=.002) of expressing a lack of confidence. Individuals with good self-rated health (OR 0.41, 95% CI 0.21-0.81, *P*=.01), who had a usual source of care (OR 0.49, 95% CI 0.24-1.00, *P*=.049), or who had higher levels of civic engagement (OR 0.44, 95% CI 0.20-0.97, *P*=.04) had lower odds of saying they were not confident filling out online forms. Among foreign-born Latinos, lack of English proficiency (OR 2.56, 95% CI 1.89-3.47, *P*=.001) and exclusive use of foreign language media (OR 1.49, 95% CI 1.10-2.02, *P*=.01) was related to lack of confidence filling out forms. Individuals who were near poor (OR 1.73, 95% CI 1.73-2.56, *P*=.006) or who lacked a high school education (OR 2.06, 95% CI 1.34-3.16, *P*=.001) had higher odds of not being confident filling out forms. Foreign-born Latinos who ate at least two servings of fast food per week or who had higher levels of perceptions of trust and safety in their neighborhood were less likely to express a lack of confidence with filling out online forms (OR 0.69, 95% CI 0.55-0.87, *P*=.002).

## Discussion

### Overview

Online interventions are viewed as a low-cost platform to deliver health information and interventions [38,39], but our findings show Latinos, particularly foreign-born Latinos, may not benefit from this shift to online delivery of health-related content and care. Our findings regarding a gap in Internet use and online health information-seeking behavior between non-Hispanic whites and Latinos are consistent with previous literature showing a disparity in Internet usage for online health information-seeking behavior in the general US population [15,18,19]. Our analysis adds to this literature in a number of ways. First, we show that there is a gap in reported confidence to fill out Internet forms between Latinos and non-Hispanic whites because Latinos have higher odds of not being confident in their ability to fill out online forms. Our study also contributes to the literature by showing a gap in online health information-seeking behavior and a confidence in filling out online forms between US- and foreign-born Latinos. The lower level of form-related confidence between Latinos and non-Hispanic whites may serve as a barrier to accessing health-related information through electronic records and in other important contexts. Improving confidence to fill out online forms may help in bridging this knowledge gap that Latinos have displayed in a number of contexts where filling out online forms allows them to participate in further information-seeking behavior. For example, many online smoking cessation and other forums require participants to provide health information and fill out forms. Feeling less confident or having difficulty with this type of task may limit access to these and other forms of online health information, such as medical records. Feeling less confident or having difficulty with online forms may also hinder the collection of health-related data on Latinos because they may be less inclined to participate in online data collection venues.

### Principal Findings

Education has been shown to be a social determinant of online health information-seeking behavior among the general US and the Latino population [26,40], but findings regarding age and gender have been mixed [11,15,26,27]. Our findings show that, consistent with studies on the general US population, education is related to online health information-seeking behavior [40]. Although one study found being male was related to online health information-seeking behavior among Latinos [26], our findings are consistent with other studies showing being female is associated with online health information-seeking behavior among the general US population, and Latinos in particular. We found that only those aged 65 or older had significantly lower odds of using the Internet compared to individuals aged 18 to 24 years. Studies of the US general population show individuals aged 65 or older are less likely to engage in online health information-seeking behavior [40]. A prior study showed language was a significant predictor of online health information-seeking behavior among Latinos [15], our findings show that this was a significant determinant among native-born individuals only. When considering confidence in filling out online forms, we found that low education was associated with lower levels of form-related confidence. Additionally, language was a significant predictor of confidence for foreign-born, but not US-born, Latinos.

Increasing digital literacy among Latinos in California (and elsewhere) should be a priority because these individuals are more likely to be never users or discontinued Internet users compared to non-Hispanic whites. They also have low levels of Internet efficacy, even when controlling for primarily speaking Spanish at home [19], although recent studies have shown that Internet access and online health information-seeking behavior among foreign-born Latinos may be increasing [28]. Our findings show that although males and those with lower education need to be targeted with education campaigns among the general Latino population [26], the correlates of online health information-seeking behavior varied by nativity. For example, a lack of English proficiency is significantly related to online



XSL•FO
RenderX

health information-seeking behavior among US-born individuals, but related to lack of form confidence among foreign-born individuals. Previous studies have shown that in the general US population, those who have difficulty accessing care are more likely to access online health information-seeking behavior [41], but we found that US-born Latinos who did not have a usual source of care had lower odds of engaging in online health information-seeking behavior. There was no relationship between access to care and online health information-seeking behavior among foreign-born Latinos. Additionally, there appeared to be an association, although it was not statistically significant, with having a chronic disease and with online health information-seeking behavior among native-born individuals ($P$=.06), but not among foreign-born individuals ($P$=.24). Because chronic diseases require significant support, knowledge, and self-care, individuals with chronic diseases should be targets of campaigns to increase online health information-seeking behavior among foreign-born Latinos. When we examined individuals who engaged in health-risk behaviors, we found that only smoking and binge drinking were related to online health information-seeking behavior among all Latinos. However, we found that this association varied by nativity. Among native-born Latinos, binge drinking was associated with higher odds of online health information-seeking behavior, whereas among foreign-born individuals, smokers had lower odds of engaging in online health information-seeking behavior.

### Limitations

This study suffers from the limitations of cross-sectional self-reported data; causality cannot be determined and responses may be biased by the limitations of memory. Social desirability may also have biased responses, particularly when reporting characteristics such as weight and substance use. Additionally, general literacy, health literacy, computer literacy, and eHealth literacy—all of which influence online health information-seeking behavior—were not assessed by the CHIS. This sample is not a general US sample, but is limited to individuals in California. As a result, findings are not generalizable to the US population, but may suggest trends taking place in states with large Latino populations, particularly in the West and in areas with growing populations of Latinos.

### Conclusion

Latinos, particularly first foreign-born individuals, are at an increased risk of being left behind as the move to increase online content delivery and care expands. When considering Latinos, it is important to note that there are nativity differences in the correlates of online health information-seeking behavior. Our research also identifies a significant gap in confidence regarding filling out online forms between individuals who are first US- and foreign-born, whose English proficiency is low and who only consume media in non-English languages, particularly among foreign-born Latinos. This indicates that as health information and online health portals become more popular, education and training for foreign-born individuals, and online health portals that are in Spanish and use Spanish forms, should be considered. Additionally, education was significantly related to a lack of confidence in filling out online forms, indicating that perhaps usability and simplicity should be a priority for online sites. This may also indicate that until digital literacy can be increased among all groups, data may need to be gathered in offline formats.

There are some groups within our sample that engage in high-risk health-related behaviors. Those US-born Latinos who binge drink are less likely to search for health information-seeking behavior online, suggesting that alternate channels should be used to supply this population with health-related information. It may be possible to provide health messages in bars or other places where binge drinkers are likely to frequent. This has already been done in the context of tobacco control. Further, foreign-born Latinos who eat two or more servings of fast food per week were less likely to be confident in filling out online forms. This may suggest that in some communities in California, particularly those that contain a small number of fast food establishments that could be tracked by researchers, researchers could pilot an eHealth literacy intervention to take place in these establishments.

### Authors' Contributions

MG designed the initial analysis and wrote the first draft with ASJ. JE and ASJ helped to refine the analysis and to edit subsequent drafts. All authors approved the interpretation of the analysis, helped to revise the manuscript, and gave final approval of the version to be published.

### Conflicts of Interest

None declared.

### References

1. [No authors listed]. Drug company's consumer health portal encourages return visits. Internet Healthc Strateg 2004 Jul;6(7):7-8. [Medline: 15346971]
2. Zickuhr K. Generations 2010. Washington, DC: Pew Internet and American Life Project; 2010 Dec 16. URL: http://www.pewinternet.org/files/old-media//Files/Reports/2010/PIP_Generations_and_Tech10.pdf [accessed 2016-03-30] [WebCite Cache ID 6gNkz2cVw]
3. Fox S, Duggan M. Health Online 2013. Washington, DC: Pew Internet and American Life Project; 2013 Jan 15. URL: http://www.pewinternet.org/files/old-media//Files/Reports/PIP_HealthOnline.pdf [accessed 2016-03-29] [WebCite Cache ID 6gH7JAWS2]



XSL•FO
RenderX

4. Pechmann C, Pan L, Delucchi K, Lakon CM, Prochaska JJ. Development of a Twitter-based intervention for smoking cessation that encourages high-quality social media interactions via automessages. J Med Internet Res 2015;17(2):e50 [FREE Full text] [doi: 10.2196/jmir.3772] [Medline: 25707037]
5. Shahab L, Brown J, Gardner B, Smith SG. Seeking health information and support online: does it differ as a function of engagement in risky health behaviors? Evidence from the health information national trends survey. J Med Internet Res 2014;16(11):e253 [FREE Full text] [doi: 10.2196/jmir.3368] [Medline: 25380308]
6. Kruse CS, Bolton K, Freriks G. The effect of patient portals on quality outcomes and its implications to meaningful use: a systematic review. J Med Internet Res 2015;17(2):e44 [FREE Full text] [doi: 10.2196/jmir.3171] [Medline: 25669240]
7. Hale T, Cotten S, Drentea P, Goldner M. Rural-urban differences in general and health-related Internet use. Am Behav Sci 2010 Mar 09;53(9):1304-1325. [doi: 10.1177/0002764210361685]
8. Atkinson NL, Saperstein SL, Pleis J. Using the internet for health-related activities: findings from a national probability sample. J Med Internet Res 2009;11(1):e4 [FREE Full text] [doi: 10.2196/jmir.1035] [Medline: 19275980]
9. Lopez M, Gonazalez-Barrera A, Patten E. Closing the Digital Divide: Latinos and Technology Adoption. Washington, DC: Pew Hispanic Center; 2013 Mar 07. URL: http://www.pewhispanic.org/files/2013/03/Latinos_Social_Media_and_Mobile_Tech_03-2013_final.pdf [accessed 2016-03-28] [WebCite Cache ID 6gNl80Aqy]
10. Laz TH, Berenson AB. Racial and ethnic disparities in internet use for seeking health information among young women. J Health Commun 2013;18(2):250-260 [FREE Full text] [doi: 10.1080/10810730.2012.707292] [Medline: 23130608]
11. Peña-Purcell N. Hispanics' use of Internet health information: an exploratory study. J Med Libr Assoc 2008 Apr;96(2):101-107 [FREE Full text] [doi: 10.3163/1536-5050.96.2.101] [Medline: 18379664]
12. Sarkar U, Karter AJ, Liu JY, Adler NE, Nguyen R, López A, et al. Social disparities in internet patient portal use in diabetes: evidence that the digital divide extends beyond access. J Am Med Inform Assoc 2011 May 1;18(3):318-321 [FREE Full text] [doi: 10.1136/jamia.2010.006015] [Medline: 21262921]
13. Sligo F, Jameson A. The knowledge-behavior gap in the use of online health information. J Assoc Inf Sci Tech 2000;51(9):858-869 DOI: 10.1002/(SICI)1097-4571(2000)51:9(858::AID-ASI80)3.0.CO;2-Q.
14. Neter E, Brainin E. eHealth literacy: extending the digital divide to the realm of health information. J Med Internet Res 2012 Jan;14(1):e19 [FREE Full text] [doi: 10.2196/jmir.1619] [Medline: 22357448]
15. De Jesus M, Xiao C. Predicting Internet use as a source of health information: a “language divide” among the Hispanic population in the United States. Policy Internet 2012;4(2):1-11. [doi: 10.1515/1944-2866.1178]
16. Norman CD, Skinner HA. eHealth literacy: essential skills for consumer health in a networked world. J Med Internet Res 2006 Jun;8(2):e9 [FREE Full text] [doi: 10.2196/jmir.8.2.e9] [Medline: 16867972]
17. Krogstad J, Stepler R, Lopez M. English Proficiency on the Rise among Latinos: US Born Driving Language Changes. Washington, DC: Pew Research Center; 2015 May 12. URL: http://www.pewhispanic.org/files/2015/05/2015-05-12_hispanics-english-proficiency_FINAL.pdf [accessed 2016-03-28] [WebCite Cache ID 6gNkd7cek]
18. Cunningham PJ, Hibbard J, Gibbons CB. Raising low 'patient activation' rates among Hispanic immigrants may equal expanded coverage in reducing access disparities. Health Aff (Millwood) 2011 Oct;30(10):1888-1894 [FREE Full text] [doi: 10.1377/hlthaff.2009.0805] [Medline: 21976331]
19. Choi NG, Dinitto DM. The digital divide among low-income homebound older adults: Internet use patterns, eHealth literacy, and attitudes toward computer/Internet use. J Med Internet Res 2013 May;15(5):e93 [FREE Full text] [doi: 10.2196/jmir.2645] [Medline: 23639979]
20. Demark-Wahnefried W, McClelland J, Campbell MK, Hoben K, Lashley J, Graves C, et al. Awareness of cancer-related programs and services among rural African Americans. J Natl Med Assoc 1998 Apr;90(4):197-202. [Medline: 9581439]
21. Ward JA, Baum S, Ter MJ, Thomsen CA, Maibach EW. The value and impact of the Cancer Information Service telephone service. Part 4. J Health Commun 1998;3 Suppl:50-70. [doi: 10.1080/108107398127256] [Medline: 10977270]
22. Squiers L, Bright MA, Rutten LJ, Atienza AA, Treiman K, Moser RP, et al. Awareness of the National Cancer Institute's Cancer Information Service: results from the Health Information National Trends Survey (HINTS). J Health Commun 2006;11 Suppl 1:117-133. [doi: 10.1080/10810730600637517] [Medline: 16641078]
23. Second-Generation Americans: A Portrait of Adult Children Immigrants. Washington, DC: Pew Research Center; 2013 Feb 07. URL: http://www.pewsocialtrends.org/files/2013/02/FINAL_immigrant_generations_report_2-7-13.pdf [accessed 2016-03-30] [WebCite Cache ID 6gNlzno1C]
24. Between Two Worlds: How Young Latinos Come of Age in America. Washington, DC: Pew Hispanic Center; 2009 Dec 11. URL: http://www.pewhispanic.org/files/reports/117.pdf [accessed 2016-03-29] [WebCite Cache ID 6gNmCXWS7]
25. Finney Rutten LJ, Hesse BW, Moser RP, Ortiz MA, Kornfeld J, Vanderpool RC, et al. Socioeconomic and geographic disparities in health information seeking and Internet use in Puerto Rico. J Med Internet Res 2012;14(4):e104 [FREE Full text] [doi: 10.2196/jmir.2007] [Medline: 22849971]
26. Lee YJ, Boden-Albala B, Larson E, Wilcox A, Bakken S. Online health information seeking behaviors of Hispanics in New York City: a community-based cross-sectional study. J Med Internet Res 2014;16(7):e176 [FREE Full text] [doi: 10.2196/jmir.3499] [Medline: 25092120]
27. Lee YJ, Boden-Albala B, Quarles L, Wilcox A, Bakken S. Predictors of health information-seeking behaviors in hispanics. NI 2012 (2012) 2012;2012:243 [FREE Full text] [Medline: 24199094]




XSL•FO
RenderX

28. Selsky C, Luta G, Noone A, Huerta EE, Mandelblatt JS. Internet access and online cancer information seeking among Latino immigrants from safety net clinics. J Health Commun 2013;18(1):58-70 [FREE Full text] [doi: 10.1080/10810730.2012.688248] [Medline: 23066874]
29. Livingston G, Minushkin S, Cohn D. Hispanics and Health Care in the United States: Access, Information and Knowledge. A Joint Pew Hispanic Center and Robert Wood Johnson Foundation Research Report. Washington, DC: Pew Hispanic Center & Robert Wood Johnson Foundation; 2008. URL: http://www.pewhispanic.org/files/reports/91.pdf [accessed 2016-03-30] [WebCite Cache ID 6gNml9A0y]
30. Gordon NP, Iribarren C. Health-related characteristics and preferred methods of receiving health education according to dominant language among Latinos aged 25 to 64 in a large Northern California health plan. BMC Public Health 2008;8:305 [FREE Full text] [doi: 10.1186/1471-2458-8-305] [Medline: 18782454]
31. United States Census Bureau. QuickFacts: United States URL: http://quickfacts.census.gov/qfd/states/06000.html [accessed 2015-08-23] [WebCite Cache ID 6b0Pgr4J1]
32. Miyawaki CE. Caregiving practice patterns of Asian, Hispanic, and non-Hispanic white American family caregivers of older adults across generations. J Cross Cult Gerontol 2016 Mar;31(1):35-55. [doi: 10.1007/s10823-016-9281-5] [Medline: 26810575]
33. Kim JE, Tsoh JY. Cigarette smoking among socioeconomically disadvantaged young adults in association with food insecurity and other factors. Prev Chronic Dis 2016;13:E08 [FREE Full text] [doi: 10.5888/pcd13.150458] [Medline: 26766849]
34. Morimoto LM, McCauley K, Ma X, Wiemels JL, Chokkalingam AP, Metayer C. Birth weight, fetal growth, and risk of pediatric rhabdomyosarcoma: an updated record linkage study in California. Ann Epidemiol 2016 Feb;26(2):141-145. [doi: 10.1016/j.annepidem.2015.11.007] [Medline: 26795698]
35. Pew Hispanic Center Hispanic Trends. Demographic profile of Hispanics in California, 2011 URL: http://www.pewhispanic.org/states/state/ca/ [accessed 2016-03-30] [WebCite Cache ID 6gNmusNwA]
36. Stepler R, Brown A. Summary Statistical Portrait of Hispanics in the United States, 1980–2013. Washington, DC: Pew Research Center; 2015. URL: http://www.pewhispanic.org/files/2015/05/2015-05-12_statistical-portrait-of-hispanics-in-the-united-states-2013_final.pdf [accessed 2016-03-29] [WebCite Cache ID 6gNmbJang]
37. UCLA Center for Health Policy Research. 2014 Jun. California Health Interview Survey 2011-12: public use data files URL: http://healthpolicy.ucla.edu/chis/data/Pages/public-use-data.aspx [accessed 2014-07-04] [WebCite Cache ID 6i7YuIOYo]
38. Griffiths F, Lindenmeyer A, Powell J, Lowe P, Thorogood M. Why are health care interventions delivered over the internet? A systematic review of the published literature. J Med Internet Res 2006;8(2):e10 [FREE Full text] [doi: 10.2196/jmir.8.2.e10] [Medline: 16867965]
39. Grey M, Liberti L, Whittemore R. Costs of development and maintenance of an Internet program for teens with type 1 diabetes. Health Technol (Berl) 2015 Jul;5(2):127-133. [doi: 10.1007/s12553-015-0109-z] [Medline: 26213677]
40. Bundorf MK, Wagner TH, Singer SJ, Baker LC. Who searches the internet for health information? Health Serv Res 2006 Jun;41(3 Pt 1):819-836 [FREE Full text] [doi: 10.1111/j.1475-6773.2006.00510.x] [Medline: 16704514]
41. Amante DJ, Hogan TP, Pagoto SL, English TM, Lapane KL. Access to care and use of the Internet to search for health information: results from the US National Health Interview Survey. J Med Internet Res 2015;17(4):e106 [FREE Full text] [doi: 10.2196/jmir.4126] [Medline: 25925943]


## Abbreviations

**BMI:** body mass index
**CHIS:** California Health Interview Survey
**FPL:** federal poverty level


*Edited by G Eysenbach; submitted 27.08.15; peer-reviewed by S Paige, J Li; comments to author 21.01.16; revised version received 26.02.16; accepted 21.03.16; published 04.07.16.*

*Please cite as:*
*Gonzalez M, Sanders-Jackson A, Emory J*
*Online Health Information-Seeking Behavior and Confidence in Filling Out Online Forms Among Latinos: A Cross-Sectional Analysis of the California Health Interview Survey, 2011-2012*
*J Med Internet Res 2016;18(7):e184*
*URL: http://www.jmir.org/2016/7/e184/*
*doi:10.2196/jmir.5065*
*PMID:27377466*




XSL•FO
RenderX

©Mariaelena Gonzalez, Ashley Sanders-Jackson, Jason Emory. Originally published in the Journal of Medical Internet Research (http://www.jmir.org), 04.07.2016. This is an open-access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/2.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on http://www.jmir.org/, as well as this copyright and license information must be included.



XSL•FO
RenderX

Original Paper

# Modelling and Predicting eHealth Usage in Europe: A Multidimensional Approach From an Online Survey of 13,000 European Union Internet Users

Joan Torrent-Sellens[1*], PhD; Ángel Díaz-Chao[2*], PhD; Ivan Soler-Ramos[1], MS(c); Francesc Saigí-Rubió[3*], PhD

[1]Department of Economics and Business, Universitat Oberta de Catalunya, Barcelona, Spain

[2]Applied Economics Department, Rey Juan Carlos University, Madrid, Spain

[3]Department of Health Science, Universitat Oberta de Catalunya, Barcelona, Spain

[*]these authors contributed equally

**Corresponding Author:**
Francesc Saigí-Rubió, PhD
Department of Health Science
Universitat Oberta de Catalunya
Av. Tibidabo 39-43
Barcelona, 08035
Spain
Phone: 34 34 933263622
Fax: 34 934505201
Email: fsaigi@uoc.edu

## Abstract

**Background:** More advanced methods and models are needed to evaluate the participation of patients and citizens in the shared health care model that eHealth proposes.

**Objective:** The goal of our study was to design and evaluate a predictive multidimensional model of eHealth usage.

**Methods:** We used 2011 survey data from a sample of 13,000 European citizens aged 16–74 years who had used the Internet in the previous 3 months. We proposed and tested an eHealth usage composite indicator through 2-stage structural equation modelling with latent variables and measurement errors. Logistic regression (odds ratios, ORs) to model the predictors of eHealth usage was calculated using health status and sociodemographic independent variables.

**Results:** The dimensions with more explanatory power of eHealth usage were health Internet attitudes, information health Internet usage, empowerment of health Internet users, and the usefulness of health Internet usage. Some 52.39% (6811/13,000) of European Internet users' eHealth usage was more intensive (greater than the mean). Users with long-term health problems or illnesses (OR 1.20, 95% CI 1.12–1.29) or receiving long-term treatment (OR 1.11, 95% CI 1.03–1.20), having family members with long-term health problems or illnesses (OR 1.44, 95% CI 1.34–1.55), or undertaking care activities for other people (OR 1.58, 95% CI 1.40–1.77) had a high propensity toward intensive eHealth usage. Sociodemographic predictors showed that Internet users who were female (OR 1.23, 95% CI 1.14–1.31), aged 25–54 years (OR 1.12, 95% CI 1.05–1.21), living in larger households (3 members: OR 1.25, 95% CI 1.15–1.36; 5 members: OR 1.13, 95% CI 0.97–1.28; ≥6 members: OR 1.31, 95% CI 1.10–1.57), had more children <16 years of age (1 child: OR 1.29, 95% CI 1.18–1.14; 2 children: OR 1.05, 95% CI 0.94–1.17; 4 children: OR 1.35, 95% CI 0.88–2.08), and had more family members >65 years of age (1 member: OR 1.33, 95% CI 1.18–1.50; ≥4 members: OR 1.82, 95% CI 0.54–6.03) had a greater propensity toward intensive eHealth usage. Likewise, users residing in densely populated areas, such as cities and large towns (OR 1.17, 95% CI 1.09–1.25), also had a greater propensity toward intensive eHealth usage. Educational levels presented an inverted U shape in relation to intensive eHealth usage, with greater propensities among those with a secondary education (OR 1.08, 95% CI 1.01–1.16). Finally, occupational categories and net monthly income data suggest a higher propensity among the employed or self-employed (OR 1.07, 95% CI 0.99–1.15) and among the minimum wage stratum, earning ≤€1000 per month (OR 1.66, 95% CI 1.48–1.87).

**Conclusions:** We provide new evidence of inequalities that explain intensive eHealth usage. The results highlight the need to develop more specific eHealth practices to address different realities.

***(J Med Internet Res 2016;18(7):e188)*** doi:10.2196/jmir.5605



XSL•FO
RenderX

**KEYWORDS**

Internet; eHealth usage; health care; health drivers; health barriers; health attitude; health information; health empowerment; information and communication technologies; structural equation modelling; Europe

## Introduction

In recent years, there has been considerable development in the field of eHealth services. With eHealth, a wide range of new opportunities has emerged to improve people's health status through the use of information and communication technologies (ICTs) in general and the Internet in particular [1-3]. In the current context of severe constraints on health budgets, eHealth is becoming a very useful instrument to improve equality of access to, and the quality of, health care [4]. However, despite being widely used and having different characteristics depending on its application, eHealth has not been precisely defined. It is an emergent practice at the intersection of medical informatics, public health, and business [3]. In the face of this conceptual limitation, several important contributions have been made in the literature. Oh et al [5] compared 51 definitions of eHealth, and van Gemert-Pijnen et al [6] identified 16 eHealth frameworks based on their theoretical antecedents, their different visions, and the strategies or principles for increasing the uptake and impact of eHealth technologies. However, the most commonly cited definition on the Internet is Eysenbach's [3] and it constituted the starting point of our study.

With new developments in wireless technologies, Web 2.0, and Media 3.0, eHealth has continued to profoundly change health care, which is shifting from an individual approach (care of acute health problems) toward a population approach (disease prevention and management through online communities) [7]. Consequently, health care provision models are evolving in a way that empowers patients to take care of, and make decisions on, their health [7]. Access to a wide range of health information, which used to be hard for the general public to obtain [8,9], and the sharing and posting of user content or comments in blogs and videos [10] have also been identified as means to enable greater patient empowerment and better self-care [11]. Today, patient-centered health care is recognized as the cornerstone of health care systems because it allows for improvements in health care outcomes and quality [12] to be made by reducing costs [13] and resource usage [14]. More and more patients are now better prepared for (they have the necessary knowledge to make decisions) and more informed about a wide range of health care-related topics [15-17]. They want to use ICTs in general and the Internet in particular to communicate with each other and share personal information about health [18,19].

In the context and objectives of the digital agenda for Europe, the eHealth Action Plan 2012–2020 promotes patient-centered care, thereby empowering citizens to make health decisions [20]. The aim is to foster the sustainability and efficiency of European health care systems by unlocking innovation and promoting changes in health care organizations. However, there is still very little consensus on exactly what the implications are of getting patients and citizens involved in this shared health care model [21,22], on how eHealth technologies match users' anticipated needs [18,23,24], or, indeed, on what the main indicators of participation should be or how they should be measured [25]. It is therefore very difficult to compare the results obtained [26,27]. Obtaining empirical evidence of inequalities in health Internet usage is a work in progress [26,28,29], and not all studies consider the necessary variables [28,30,31] or are suitably adapted to factors that could foster health Internet usage in a constantly changing digital environment [26,27]. In addition, research on health Internet usage as a whole is still very scant in Europe [32] because most of the literature comes from the United States.

Since any impact fluctuates over time and in a given context [33-35], it has been suggested in the literature that there is a need to use more advanced methods to evaluate the participation of patients and citizens in this shared health care model. Social theory [36] points out that the analysis of health Internet usage disparities requires a more integrated approach that takes into account the drivers and barriers presented by the characteristics of people, of socioeconomic and cultural environments, and of technology usage [37,38]. Among other dimensions having an impact on health Internet usage, the sex and age of patients and citizens [27,39,40] have been noted, as have sociodemographic factors such as education or literacy [41,42], health status [28,39,40,43,44], and psychographic indicators such as the trust that people place in the Internet, in their own physicians, or in the health care system. Only a comprehensive examination of these dimensions will facilitate a better understanding of the complexity of citizens' and patients' eHealth usage [6]. Indeed, citizens' and patients' lack of knowledge of eHealth-related opportunities and challenges has already been identified in the eHealth Action Plan 2012–2020 as the main barrier to the acceptance of eHealth solutions in Europe [20].

Thus, the main aim of this work was to model and predict eHealth usage in Europe. We designed a multidimensional model for this purpose. The model has 9 dimensions and 88 indicators. We constructed an eHealth usage composite indicator by means of a structural equation modelling (SEM) analysis of a sample of 13,000 European Internet users in 2011. We then conducted a study to establish the indicator's main predictors, especially the Internet users' sociodemographic variables and health status. The results obtained provide new evidence of eHealth usage in Europe and have implications for the design of public health policies.

## Methods

### Participants and Procedure

Data for this study were drawn from the Strategic Intelligence Monitor on Personal Health Systems Phase 2 (SIMPHS2) research project "Citizens and ICT for health in 14 European countries: results from an online panel" [45]. The study was carried out by the Institute for Prospective Technological Studies in cooperation with the European Commission Directorate General for Information Society and Media, now the Directorate General for Communications Networks, Content and



XSL•FO
RenderX

Technology. The SIMPHS2 citizen panel survey's analysis of user demand had as its main objectives (1) to develop typologies of digital health care users and measure the impact of ICT and the Internet on health status, health care demand, and health management, and (2) to identify factors that can enhance or inhibit the role and use of personal health systems from a citizen's perspective with special emphasis on mHealth, remote patient monitoring and treatment disease management, telecare, telemedicine, and wellness [45].

Our study used survey data for a sample of 13,000 European citizens aged 16–74 years who had used the Internet in the previous 3 months (Multimedia Appendix 1). The sampling universe comprised 171,859,356 European citizens aged 16–74 years with an overall margin of error of ±0.88 in the case of maximum indetermination p=q=50%, for a confidence level of 95.5%. The sample had two essential characteristics. First, we chose an equal-sized sample for each of the 13 countries being studied, that is to say, 1000 interviews for each country in the sample: Austria, Belgium, Denmark, Estonia, Finland, France, Germany, Italy, the Netherlands, Sweden, Slovakia, Slovenia, and Spain (public data are available for 13 countries). The country-specific margin of error was ±3.16 in the case of maximum indetermination p=q=50%, for a confidence level of 95.5%. Second, we chose to use a fully representative sample for the distribution of the target population, according to sex and age group. The demographic groups are organized by the cross-referenced quotas of sex and age group, as follows: women aged 16–24 years (±2.78), women aged 25–54 years (±1.58), women aged 55–74 years (±3.08), men aged 16–24 years (±2.73), men aged 25–54 years (±1.56), and men aged 55–74 years (±2.89).

The questionnaire used in the survey contained 47 questions grouped into 5 dimensions (Multimedia Appendix 2): (1) health status, and health care and social care services use (12 questions), (2) health attitude and health information sources (5 questions) (3) Internet and ICT uses (2 questions), (4) health-related use of ICTs and the Internet (15 questions), and (5) sociodemographic profile of participants (13 questions). The survey was answered by European Internet users in online interviews, lasting for half an hour each, and in a native language of each country. A study presentation paragraph was written to inform potential respondents about the confidentiality of any data provided and the academic aim of the research. The European Internet users voluntarily answered the questionnaire and did not receive any payment in cash or kind. While the questionnaire was being implemented, an expert was on hand at all times (by email) to resolve any queries that the respondents had. The respondent citizens were selected by means of probability sampling applied to each country universe. The net response rate was 20.72%. To achieve 13,000 responses, it was necessary to send 65,126 invitations, to which 19,731 responses were received. Of the responses received, 6731 were excluded, either because they did not fall into the required quotas for Internet use (6236) or because they had been rejected (495). The reasons for rejecting a questionnaire were either that they were incomplete or that the consistency of responses was poor. The fieldwork period ran from the July 20 to August 20, 2011.

The SIMPHS2 research project followed the Checklist for Reporting Results of Internet E-Surveys criteria [46]. For a more detailed explanation, see the SIMPHS2 research report [45].

## Data Analysis and Models

From an empirical perspective, explanatory factors determining eHealth usage raise two particular difficulties. First, the approach to the concept requires a multidimensional basis that is not usually captured in a single variable. In fact, the most common approaches found in the literature perform partial analyses of its various dimensions. This type of analysis has the disadvantage of not taking a full snapshot of the explanatory factors, which gives rise to the second difficulty: statistical modelling. In other words, eHealth usage can be interpreted as a latent, nonobservable concept, which therefore calls for statistical techniques that allow variables of this type, which are not directly measurable, to be used [47,48].

In the empirical literature, SEM with latent variables has been used to overcome this problem. A general SEM is a formal mathematical model. It is a set of linear equations that encompasses various types of models, such as regression analysis models, simultaneous equation systems, factor analysis, and path analysis. The main advantage of this method of analysis is the incorporation of different types of variables into the SEM. Directly observable and measurable variables, and theoretical or latent variables representing concepts that are not directly observed can therefore be incorporated. When the variable to be explained (dependent) is latent, it must be continuous, whereas dependent observed variables can be continuous, censored, binary, ordered, or categorical (ordinals), or combinations of any of these variable types [49].

This method of analysis allows us to define eHealth usage as a latent variable, thus enabling us to calculate the specific explanatory effect of the variables that it comprises. Hence, besides building an overall explanatory model of the determinants of eHealth usage, it is also possible to identify which of its explanatory dimensions are more important. In addition, SEM enables the relationships between the different observable variables included in the model (indirect effects) to be estimated. In this initial approach, however, only the direct effects are presented, that is to say, the coefficients of causality between the individual indicators and their latent dimensions, and later between the estimated dimensions and the latent variable (eHealth usage). In this context, and in order to capture the factors that explain eHealth usage in a large sample of European Internet users, we proposed and tested a 2-stage SEM with latent variables and measurement errors for 2011.

We applied the 2-stage empirical estimation methodology as follows: in the first stage, we tested the causal relationships among 88 indicators and the 9 latent dimensions describing eHealth usage in Europe, and in the second stage, we tested the causal relationships among the indicators constructed for those 9 dimensions (based on the coefficients from the first stage) and the latent construct of eHealth usage. Finally, after applying the coefficients obtained from the second stage, we constructed an eHealth usage indicator and determined its mean values (total and for the 9 dimensions). This methodology involved the design



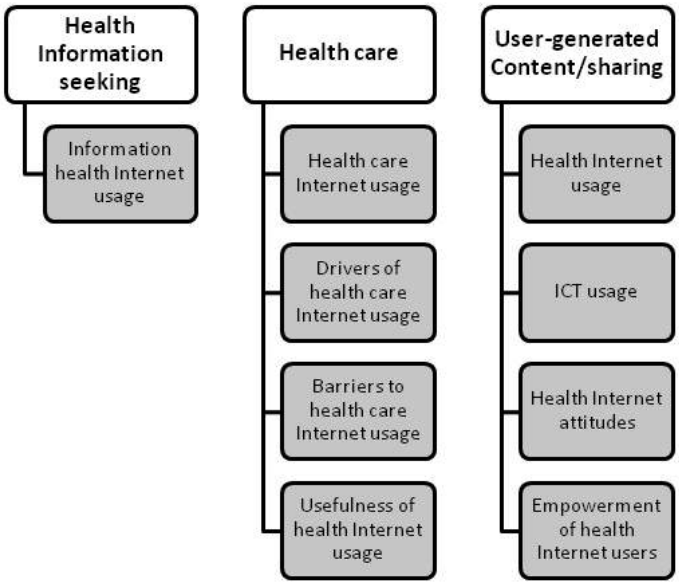
XSL•FO
RenderX

and statistical testing of 10 empirical models: 9 models for the first stage and 1 model for the second stage.

Several eHealth definitions highlight growing patient empowerment (access to information and ability to use it) and point to the potential of eHealth to facilitate doctor-patient communication, partnership, and shared decision making [3,41,50]. Figure 1 shows the multidimensional model of eHealth usage with 9 dimensions grouped in 3 domains relevant to health usage: health information seeking, health care, and user-generated content and sharing. The 9 explanatory model dimensions and variables are as follows: *dimension 1: health Internet usage*, captured by a set of 14 variables measuring the frequency of usage (Multimedia Appendix 3); *dimension 2: health care Internet usage*, captured by a set of 10 variables measuring the frequency of usage (Multimedia Appendix 4); *dimension 3: drivers of health care Internet usage*, captured by a set of 8 variables measuring the factors that Internet users consider relevant when evaluating an Internet health site (Multimedia Appendix 5); *dimension 4: barriers to health care Internet usage*, captured by a set of 10 variables measuring the factors that Internet users regarded as barriers when evaluating Internet health care (Multimedia Appendix 6); *dimension 5: usefulness of health Internet usage*, captured by a set of 13 variables measuring the Internet user's perceived usefulness of health Internet usage (Multimedia Appendix 7); *dimension 6: ICT usage*, captured by a set of 15 variables measuring the frequency of usage (Multimedia Appendix 8); *dimension 7: information health Internet usage*, captured by a set of 7 variables measuring the Internet user's perceived judgment of information health usage (Multimedia Appendix 9); *dimension 8: health Internet attitudes*, captured by a set of 6 variables measuring the Internet user's perceived feelings about health Internet usage (Multimedia Appendix 10); *dimension 9: empowerment of health Internet users*, captured by a set of 5 variables measuring the Internet user's perceived results of health Internet usage (Multimedia Appendix 11).

Additionally, we calculated the logistic regression to model the predictors of eHealth usage using health status and sociodemographic independent variables. For each independent variable, we calculated odds ratios (ORs) and their 95% CIs. We used IBM SPSS Amos v.22 (IBM Corp) for all calculations.

**Figure 1.** Flow diagram detailing the multidimensional model of eHealth usage. ICT: information and communication technologies.




## Results

### eHealth Usage Composite Indicator

Table 1 shows the results (standardized coefficients and measurement errors) of the first stage of estimating the explanatory factors of eHealth usage in Europe in 2011. In this first stage, we estimated the causal relationships among 88 indicators and the 9 dimensions describing eHealth usage by using an SEM with measurement errors. First, it should be noted that all the variables specified in the model were statistically significant (99% confidence level). Second, the goodness-of-fit measurements for the 9 proposed models were highly satisfactory. Thus, the normed fit index (NFI), relative fit index (RFI), incremental fit index (IFI), Tucker-Lewis index (TLI), and comparative fit index (CFI) had very high values, approaching the optimal value of 1. The root mean square error of approximation (RMSEA) values were <0.08, thus corroborating the validity of the estimated models.



XSL•FO
RenderX

**Table 1.** Explanatory factors of eHealth usage in Europe (first stage)[a] in 2011.

| Dimension/variable | | Standardized coefficient | *P* value | Error | *P* value |
|---|---|---|---|---|---|
| **1. Health Internet usage** | | | | 0.698 | <.001 |
| 1. | Look for information about a physical illness | 0.536 | <.001 | 1.732 | <.001 |
| 2. | Look for information about wellness or lifestyle | 0.545 | <.001 | 1.955 | <.001 |
| 3. | Buy medicine or vitamins online | 0.779 | <.001 | 2.558 | <.001 |
| 4. | Participate in an online support group with people | 0.774 | <.001 | 2.761 | <.001 |
| 5. | Participate in social networking sites | 0.790 | <.001 | 2.301 | <.001 |
| 6. | Use email or Web to communicate with a doctor's office | 0.713 | <.001 | 3.301 | <.001 |
| 7. | Click on a health or medical Web's privacy policy | 0.682 | <.001 | 3.750 | <.001 |
| 8. | Describe a medical condition to get advice from an online doctor | 0.783 | <.001 | 2.645 | <.001 |
| 9. | Describe a medical condition to get advice from other online users | 0.822 | <.001 | 1.905 | <.001 |
| 10. | Bookmark or favorite a health website | 0.725 | <.001 | 2.426 | <.001 |
| 11. | Look to see what company is providing the information on a health website | 0.681 | <.001 | 2.661 | <.001 |
| 12. | Look for information about a mental health issue | 0.749 | <.001 | 2.209 | <.001 |
| 13. | Disclose medical information on social networking sites | 0.821 | <.001 | 2.329 | <.001 |
| 14. | Disclose medical information on websites to share files | 0.814 | <.001 | 2.516 | <.001 |
| Goodness-of-fit indexes: NFI[b]: 0.986; RFI[c]: 0.979; IFI[d]: 0.987; TLI[e]: 0.980; CFI[f]: 0.987; RMSEA[g]: 0.041 | | | | | |
| **2. Health care Internet usage** | | | | 1.982 | <.001 |
| 15. | Make an Internet appointment with health care professionals | 0.743 | <.001 | 1.609 | <.001 |
| 16. | Receive an email from doctor, nurse, or health care organization | 0.781 | <.001 | 1.343 | <.001 |
| 17. | Have an online consultation through videoconference with health care professionals | 0.813 | <.001 | 1.675 | <.001 |
| 18. | Receive online the results of clinical or medical test | 0.801 | <.001 | 1.484 | <.001 |
| 19. | Use medical information through an Internet provider | 0.776 | <.001 | 2.098 | <.001 |
| 20. | Use medical information through an Internet health care organization | 0.812 | <.001 | 1.656 | <.001 |
| 21. | Use a game console to play games related to health or wellness | 0.739 | <.001 | 2.056 | <.001 |
| 22. | Use a health/wellness app on mobile phone | 0.790 | <.001 | 1.643 | <.001 |
| 23. | Use electronic devices to transmit clinical or medical information | 0.758 | <.001 | 1.811 | <.001 |
| 24. | Email about health promotion or health prevention | 0.670 | <.001 | 1.906 | <.001 |
| Goodness-of-fit indexes: NFI: 0.971; RFI: 0.953; IFI: 0.971; TLI: 0.954; CFI: 0.971; RMSEA: 0.074 | | | | | |
| **3. Drivers of health care Internet usage** | | | | 0.237 | <.001 |
| 25. | Secure handling of personal information | 0.672 | <.001 | 0.287 | <.001 |
| 26. | Information in own language | 0.580 | <.001 | 0.407 | <.001 |
| 27. | Updated information | 0.737 | <.001 | 0.246 | <.001 |
| 28. | Interactivity | 0.520 | <.001 | 0.579 | <.001 |
| 29. | Involvement of health professionals | 0.867 | <.001 | 0.150 | <.001 |
| 30. | Clear statement of who is responsible for sponsoring the site | 0.586 | <.001 | 0.614 | <.001 |
| 31. | Involvement of health organizations | 0.728 | <.001 | 0.322 | <.001 |
| 32. | Involvement of governments | 0.382 | <.001 | 0.794 | <.001 |
| Goodness-of-fit indexes: NFI: 0.973; RFI: 0.934; IFI: 0.973; TLI: 0.935; CFI: 0.973; RMSEA: 0.075 | | | | | |
| **4. Barriers to health care Internet usage** | | | | 0.296 | <.001 |



XSL•FO
RenderX

| Dimension/variable | | Standardized coefficient | *P* value | Error | *P* value |
|---|---|---|---|---|---|
| 33. | Lack of digital skills | 0.583 | <.001 | 0.574 | <.001 |
| 34. | Lack of access to ICT[h] for health applications | 0.632 | <.001 | 0.452 | <.001 |
| 35. | Lack of motivation and interest | 0.666 | <.001 | 0.382 | <.001 |
| 36. | Lack of awareness | 0.730 | <.001 | 0.329 | <.001 |
| 37. | Lack of health literacy | 0.714 | <.001 | 0.352 | <.001 |
| 38. | Lack of trust | 0.832 | <.001 | 0.199 | <.001 |
| 39. | Lack of liability | 0.810 | <.001 | 0.242 | <.001 |
| 40. | Lack of privacy | 0.762 | <.001 | 0.279 | <.001 |
| 41. | Lack of security | 0.800 | <.001 | 0.232 | <.001 |
| 42. | Lack of reliability | 0.804 | <.001 | 0.219 | <.001 |
| Goodness-of-fit indexes: NFI: 0.979; RFI: 0.953; IFI: 0.980; TLI: 0.953; CFI: 0.980; RMSEA: 0.074 | | | | | |
| **5. Usefulness of health Internet usage** | | | | 0.720 | <.001 |
| 43. | ICT for health could increase other ICT uses | 0.751 | <.001 | 0.555 | <.001 |
| 44. | ICT for health could lead to greater patient satisfaction | 0.819 | <.001 | 0.372 | <.001 |
| 45. | ICT for health could improve health status | 0.782 | <.001 | 0.471 | <.001 |
| 46. | ICT for health could improve the ability to take care of one's own health | 0.816 | <.001 | 0.385 | <.001 |
| 47. | ICT for health could change behaviors toward a healthy lifestyle | 0.769 | <.001 | 0.469 | <.001 |
| 48. | ICT for health could avoid travelling expenses and time | 0.740 | <.001 | 0.567 | <.001 |
| 49. | ICT for health could improve the quality of health care services | 0.803 | <.001 | 0.407 | <.001 |
| 50. | Internet health could substitute for offline consultations with the physicians | 0.604 | <.001 | 1.022 | <.001 |
| 51. | Internet health complements offline consultations with the physicians | 0.704 | <.001 | 0.687 | <.001 |
| 52. | Quality of Internet health is aligned with the quality of offline services | 0.626 | <.001 | 0.796 | <.001 |
| 53. | Personal information could be shared with physicians through Internet due to privacy | 0.273 | <.001 | 1.202 | <.001 |
| 54. | Patients could be more comfortable with a remote monitoring system to track health | 0.626 | <.001 | 0,895 | <.001 |
| 55. | Patients could be willing to pay to access Internet health services | 0.512 | <.001 | 1.140 | <.001 |
| Goodness-of-fit indexes: NFI: 0.979; RFI: 0.953; IFI: 0.980; TLI: 0.953; CFI: 0.980; RMSEA: 0.074 | | | | | |
| **6. ICT usage** | | | | 0.037 | <.001 |
| 56. | Use a search engine to find information | 0.242 | <.001 | 0.600 | <.001 |
| 57. | Send emails with attached files | 0.344 | <.001 | 0.987 | <.001 |
| 58. | Post messages to chatrooms, newsgroups, or an online discussion forum | 0.626 | <.001 | 1.290 | <.001 |
| 59. | Use the Internet to make telephone calls | 0.520 | <.001 | 1.377 | <.001 |
| 60. | Use peer-to-peer file sharing for exchanging pictures, videos, or movies | 0.637 | <.001 | 1.042 | <.001 |
| 61. | Create a webpage | 0.552 | <.001 | 0.856 | <.001 |
| 62. | Use websites to share pictures, videos, or movies | 0.681 | <.001 | 1.103 | <.001 |
| 63. | Use a social networking site | 0.436 | <.001 | 1.972 | <.001 |
| 64. | Purchase goods or services online | 0.472 | <.001 | 0.801 | <.001 |
| 65. | Keep a blog or weblog | 0.564 | <.001 | 0.939 | <.001 |
| 66. | Use instant messaging or chat websites | 0.564 | <.001 | 1.638 | <.001 |



XSL•FO
RenderX

| Dimension/variable | | Standardized coefficient | *P* value | Error | *P* value |
|---|---|---|---|---|---|
| 67. | Do home banking | 0.184 | <.001 | 1.577 | <.001 |
| 68. | Use online software | 0.612 | <.001 | 1.230 | <.001 |
| 69. | Use the Internet through mobile phone | 0.523 | <.001 | 1.833 | <.001 |
| 70. | Use online gaming or playing games console | 0.371 | <.001 | 1.976 | <.001 |
| Goodness-of-fit indexes: NFI: 0.942; RFI: 0.912; IFI: 0.944; TLI: 0.914; CFI: 0.944; RMSEA: 0.051 | | | | | |
| **7. Information health Internet usage** | | | | 0.656 | <.001 |
| 71. | Better informed about the advice of the health care professionals | 0.792 | <.001 | 0.389 | <.001 |
| 72. | Better understanding of personal health | 0.830 | <.001 | 0.301 | <.001 |
| 73. | Better informed on what is available, so that can make own choices | 0.802 | <.001 | 0.341 | <.001 |
| 74. | Better understand the relevance of personal health | 0.817 | <.001 | 0.323 | <.001 |
| 75. | Know more about the opinions of people who are in similar situations | 0.708 | <.001 | 0.505 | <.001 |
| 76. | Better understand personal health through online discussions or experiences | 0.733 | <.001 | 0.524 | <.001 |
| 77. | Play a more active role in exchanges with health care professionals | 0.728 | <.001 | 0.547 | <.001 |
| Goodness-of-fit indexes: NFI: 0.993; RFI: 0.985; IFI: 0.993; TLI: 0.986; CFI: 0.993; RMSEA: 0.046 | | | | | |
| **8. Health Internet attitudes** | | | | 0.673 | <.001 |
| 78. | Better equipped to implement the advice of health care professionals | 0.819 | <.001 | 0.332 | <.001 |
| 79. | Better equipped to make own choices without the advice of a physician | 0.791 | <.001 | 0.433 | <.001 |
| 80. | Better equipped to make positive changes through other people | 0.805 | <.001 | 0.353 | <.001 |
| 81. | More confident in playing a more active role in relationship with physician | 0.834 | <.001 | 0.319 | <.001 |
| 82. | More confident about choices on possible treatments and solutions | 0.863 | <.001 | 0.265 | <.001 |
| 83. | More confident in discussions with the people in one's life | 0.795 | <.001 | 0.387 | <.001 |
| Goodness-of-fit indexes: NFI: 0.998; RFI: 0.991; IFI: 0.998; TLI: 0.992; CFI: 0.998; RMSEA: 0.041 | | | | | |
| **9. Empowerment health Internet users** | | | | 0.665 | <.001 |
| 84. | Make decisions on health, albeit without going against the physicians | 0.760 | <.001 | 0.486 | <.001 |
| 85. | Take a more active role in health by deciding solutions or alternative approaches | 0.840 | <.001 | 0.317 | <.001 |
| 86. | Make decisions about health on the basis of own preferences | 0.825 | <.001 | 0.384 | <.001 |
| 87. | Take a more active role in health by continuing to talk with people | 0.775 | <.001 | 0.414 | <.001 |
| 88. | Make decisions about health by relying on the experiences of other people | 0.783 | <.001 | 0.452 | <.001 |
| Goodness-of-fit indexes: NFI: 0.997; RFI: 0.988; IFI: 0.997; TLI: 0.992; CFI: 0.988; RMSEA: 0.048 | | | | | |

[a]Regression analysis: structural equation modelling; direct effects.

[b]NFI: normed fit index.

[c]RFI: relative fit index.

[d]IFI: incremental fit index.

[e]TLI: Tucker-Lewis index

[f]CFI: comparative fit index.

[g]RMSEA: root mean square error of approximation.

[h]ICT: information and communication technology.

In the health Internet usage dimension, the standardized coefficient variability is 0.3 points. The variables with the greatest explanatory power in this dimension are related to describing a medical condition to get advice from other Internet users (0.822), as well as disclosing medical information on social networking sites (0.821) or on websites (0.814). In contrast, less explanatory variables are related to finding information about physical illness (0.536) or wellness and



XSL•FO
RenderX

lifestyle (0.545). In the health care Internet usage dimension, the standardized coefficient variability is 0.14 points, between the explanatory variables related to online consultation through videoconference with health care professionals (0.813), using medical information through an Internet health care organization (0.812), and receiving emails about health promotion or health prevention (0.670). In the drivers of health care Internet usage dimension, the standardized coefficient variability is high and reaches about 0.5 points. The variable with the greatest explanatory power is the involvement of health professionals (0.867), and the variable with the least explanatory power is the involvement of governments (0.382). In the barriers to health care Internet usage dimension, variability is 0.25 points, between the lack of trust (0.832), liability (0.810), reliability (0.804), and security (0.800) and the lack of digital skills (0.583). In the usefulness of health Internet usage dimension, the explanatory variable variability is around 0.3 points, from the perceptions that ICT for health could lead to greater patient satisfaction (0.819), could improve the ability to take care of one's own health (0.816), and could improve the quality of health care services (0.803) to the willingness to pay to access Internet health services (0.512). In the ICT usage dimension, variability is the highest, and is around 0.5 points, from using the Internet to share pictures, videos, or movies (0.681), peer-to-peer file sharing (0.637), posting messages to chat rooms, newsgroups, or online discussion forums (0.626), and using online software (0.612) to using a search engine to find information (0.242) and home banking (0.184). Finally, in the information health Internet usage, health Internet attitudes, and empowerment of health Internet users dimensions, the explanatory variable variability is minimal, and all the obtained coefficients are in the range from 0.7 to 0.8 points.

Table 2 shows the results (standardized coefficients and measurement errors) of the second stage of estimating the explanatory factors of eHealth usage in Europe in 2011. In this second stage, we tested the causal relationships among the indicators constructed for the 9 dimensions describing eHealth usage (based on the coefficients from the first stage) and the latent construct of explanatory factors of eHealth usage by using an SEM with a latent dependent variable and measurement errors. First, it should be noted that all the variables specified in the model were statistically significant (95% confidence level, at least). Second, the goodness-of-fit measurements for the proposed model were highly satisfactory. Thus, the indexes NFI (0.981), RFI (0.961), IFI (0.981), TLI (0.962), and CFI (0.981) had very high values, approaching the optimal value of 1. The RMSEA value was <0.08 (0.052), thus corroborating the validity of the estimated model.

**Table 2.** Explanatory factors of eHealth usage in Europe (second stage)[a] in 2011.

| Dimension/variable | | Standardized coefficient | *P* value | Error | *P* value |
|---|---|---|---|---|---|
| **eHealth usage** | | | | 3.538 | <.001 |
| 1. | Health Internet usage | 0.099 | <.001 | 360.143 | <.001 |
| 2. | Health care Internet usage | 0.029 | <.001 | 161.145 | <.001 |
| 3. | Drivers of health care Internet usage | 0.311 | <.001 | 8.003 | <.001 |
| 4. | Barriers to health care Internet usage | 0.221 | <.001 | 21.665 | <.001 |
| 5. | Usefulness of health Internet usage | 0.547 | <.001 | 37.930 | <.001 |
| 6. | Information and communication technology usage | 0.240 | <.001 | 31.880 | <.001 |
| 7. | Information health Internet usage | 0.859 | <.001 | 5.221 | <.001 |
| 8. | Health Internet attitudes | 0.940 | <.001 | 2.146 | <.001 |
| 9. | Empowerment of health Internet users | 0.855 | <.001 | 3.446 | <.001 |
| Goodness-of-fit indexes: NFI[b]: 0.981; RFI[c]: 0.961; IFI[d]: 0.981; TLI[c]: 0.962; CFI[f]: 0.981; RMSEA[g]: 0.053 | | | | | |

[a]Regression analysis: structural equation modelling; estimated coefficients: direct effects.

[b]NFI: normed fit index.

[c]RFI: relative fit index.

[d]IFI: incremental fit index.

[e]TLI: Tucker-Lewis index

[f]CFI: comparative fit index.

[g]RMSEA: root mean square error of approximation.

The standardized coefficients obtained for the indicators of the 9 dimensions of eHealth usage in Europe highlight different explanatory capabilities. The dimensions with more-explanatory power are health Internet attitudes (0.940), information health Internet usage (0.859), empowerment of health Internet users (0.855), and usefulness of health Internet usage (0.547). ICT usage (0.240), and drivers of (0.311) and barriers to (0.221) health care Internet usage fall in the middle. Finally, the health Internet usage (0.099) and health care Internet usage (0.029) standardized coefficients have the least eHealth usage explanatory power. After applying the coefficients obtained from the second stage, we constructed an eHealth usage composite indicator and determined its mean values (Table 3).



XSL•FO
RenderX

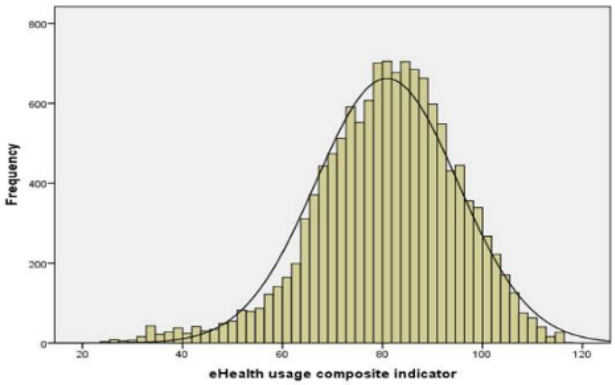
**Table 3.** eHealth usage composite indicator descriptive statistics, 2011.

| Dimension/variable | Mean | SD | Minimum | Maximum | Skewness | Kurtosis |
|---|---|---|---|---|---|---|
| 1 Health Internet usage | 25.37 | 19.07 | 10.21 | 91.93 | 1.832 | 2.705 |
| 2 Health care Internet usage | 14.50 | 12.70 | 7.68 | 69.15 | 2.768 | 7.756 |
| 3 Drivers of health care Internet usage | 16.41 | 2.98 | 5.07 | 20.29 | –1.352 | 2.530 |
| 4 Barriers to health care Internet usage | 23.21 | 4.77 | 7.33 | 29.33 | –0.963 | 1.141 |
| 5 Usefulness of health Internet usage | 28.99 | 7.37 | 8.83 | 44.13 | –0.458 | 0.375 |
| 6 Information and communication technology usage | 19.12 | 5.82 | 7.33 | 36.64 | 0.566 | –0.005 |
| 7 Information health Internet usage | 20.78 | 4.47 | 5.41 | 27.05 | –0.870 | 1.180 |
| 8 Health Internet attitudes | 18.22 | 4.29 | 4.91 | 24.54 | –0.714 | 0.756 |
| 9 Empowerment of health Internet users | 14.35 | 3.58 | 3.98 | 19.92 | –0.637 | 0.491 |
| eHealth usage composite indicator | 80.85 | 14.24 | 24.19 | 117.06 | –0.541 | 0.716 |

Figure 2 shows the histogram (frequencies and expected mean) of the values of the eHealth usage composite indicator. The mean value of this composite indicator was 80.85 points (SD 14.24, minimum to maximum range 24.19–117.06).

To capture the main predictors of eHealth usage in Europe, we performed a logistic regression using independent variables for European Internet users' health status and sociodemographic circumstances. The first step in this analysis was to recode the eHealth usage composite indicator. We therefore constructed a dichotomous eHealth usage indicator, based on the mean of the composite indicator obtained. The dichotomous eHealth usage indicator takes the value 1 when the eHealth usage composite indicator is equal to or greater than the mean, and the value 0 when less than the mean. The mean value of this dichotomous composite indicator was 0.524 points (SD 0.499, minimum to maximum range 0–1, skew –0.097, kurtosis –1.991). Some 52.39% (6811/13,000) of European Internet users' eHealth usage was more intensive (greater than the mean).

**Figure 2.** eHealth usage composite indicator histogram.




## Health Status-Related Predictors of eHealth

Table 4 shows the results of the logistic regressions (ORs) between the dichotomous eHealth usage composite indicator and the health status independent variables. We found no significant differences between European Internet users' perceived health status and more intensive eHealth usage—the variables with the highest predictive power were poor health (OR 1.30, 95% CI 1.12–1.51) and very good health (OR 1.02, 95% CI 0.94–1.11). However, the analysis of the existence of long-term health problems or illnesses did point to its predictive power. European Internet users with long-term health problems or illnesses (OR 1.20, 95% CI 1.12–1.29) or receiving long-term treatment (OR 1.11, 95% CI 1.03–1.20) had a much greater propensity toward more intensive eHealth usage. Likewise, the existence of specific health problems or illnesses determined a greater probability of more intensive eHealth usage. Specifically, these were diabetes (OR 1.01, 95% CI 0.88–1.16), stroke or cerebral hemorrhage (OR 0.95, 95% CI 0.72–1.23), cancer (OR 0.93, 95% CI 0.77–1.12), and cataract (OR 0.91, 95% CI 0.73–1.13). In contrast, users with health problems or illnesses related to chronic bronchitis and emphysema (OR 0.69, 95% CI 0.59–0.79) and osteoporosis (OR 0.63, 95% CI 0.51–0.77) had a lower propensity toward intensive eHealth usage. Finally, having family members with or caring for other people with long-term illnesses determined a greater propensity toward more intensive eHealth usage. Users with family members having long-term health problems or illnesses (OR 1.44, 95% CI 1.34–1.55) or who cared for other people with long-term health problems or illnesses (OR 1.58, 95% CI 1.40–1.77) had a greater propensity toward more intensive eHealth usage than users without such problems.



XSL•FO
RenderX

**Table 4.** Logistic regression models for odds of dichotomous eHealth usage composite indicator reporting a value of 1 (eHealth usage composite indicator greater than or equal to eHealth usage composite indicator mean) by health status, 2011.

| | OR[a] | 95% CI |
|---|---|---|
| **Perceived general health** | | |
| Very poor health | 0.91 | 0.61–1.34 |
| Poor health | 1.30 | 1.12–1.51 |
| Neither good nor poor health | 0.99 | 0.91–1.10 |
| Good health | 0.94 | 0.88–1.01 |
| Very good health | 1.02 | 0.94–1.11 |
| **Long-standing illness or health problem** | | |
| Yes | 1.20 | 1.12–1.29 |
| No | 0.83 | 0.77–0.89 |
| **Long-term medical treatment** | | |
| Yes | 1.11 | 1.03–1.20 |
| No | 0.90 | 0.84–0.97 |
| **Specific illness or health problem** | | |
| Diabetes | 1.01 | 0.88–1.16 |
| Allergy | 0.82 | 0.77–0.88 |
| Asthma | 0.87 | 0.78–0.98 |
| Hypertension | 0.86 | 0.79–0.94 |
| Long-standing muscular problem | 0.78 | 0.72–0.85 |
| Cancer | 0.93 | 0.77–1.12 |
| Cataract | 0.91 | 0.73–1.13 |
| Migraine or frequent headache | 0.83 | 0.77–0.90 |
| Chronic bronchitis, emphysema | 0.69 | 0.59–0.79 |
| Osteoporosis | 0.63 | 0.51–0.77 |
| Stroke, cerebral hemorrhage | 0.95 | 0.72–1.23 |
| Peptic, gastric, or duodenal ulcer | 0.78 | 0.68–0.91 |
| Chronic anxiety or depression | 0.72 | 0.66–0.79 |
| **Family members with long-term illness or disability** | | |
| Yes | 1.44 | 1.34–1.55 |
| No | 0.69 | 0.65–0.75 |
| **Taking care of a person with long-term illness or disability** | | |
| Yes | 1.58 | 1.40–1.77 |
| No | 0.64 | 0.57–0.71 |

[a]OR: odds ratio.

## Sociodemographic-Related Predictors of eHealth

Table 5 shows the results of the logistic regressions (ORs) between the dichotomous eHealth usage composite indicator and the sociodemographic independent variables. European Internet users who were female (OR 1.23, 95% CI 1.14–1.31) and who were aged 25–54 years (OR 1.12, 95% CI 1.05–1.21) had a greater propensity toward intensive eHealth usage than men (OR 0.82, 95% CI 0.76–0.88) or those in other age groups: 16–24 years (OR 0.97, 95% CI 0.89–1.06) and 55–74 years (OR 0.86, 95% CI 0.78–0.94). Households with more members (3 members: OR 1.25, 95% CI 1.15–1.36; 5 members: OR 1.13, 95% CI 0.97–1.28; ≥6 members: OR 1.31, 95% CI 1.10–1.57), more children <16 years of age (1 child: OR 1.29, 95% CI 1.18–1.41; 2 children: OR 1.05, 95% CI 0.94–1.17; 4 children: OR 1.35, 95% CI 0.88–2.08), and more members >65 years of age (1 member: OR 1.33, 95% CI 1.18–1.50; ≥4 members: OR 1.82, 95% CI 0.54–6.03) also had greater probabilities of more intensive eHealth usage.



XSL•FO
RenderX

**Table 5.** Logistic regressions models for odds of dichotomous eHealth usage composite indicator reporting a value of 1 (eHealth usage composite indicator greater than or equal to eHealth usage composite indicator mean) by sociodemographic conditions, 2011.

| | OR[a] | 95% CI |
|---|---|---|
| **Sex** | | |
| Male | 0.82 | 0.76–0.88 |
| Female | 1.23 | 1.14–1.31 |
| **Age range (years)** | | |
| 16–24 | 0.97 | 0.89–1.06 |
| 25–54 | 1.12 | 1.05–1.21 |
| 55–74 | 0.86 | 0.78–0.94 |
| **Number of members in the household** | | |
| 1 | 0.75 | 0.69–0.83 |
| 2 | 0.87 | 0.81–0.94 |
| 3 | 1.25 | 1.15–1.36 |
| 4 | 1.07 | 0.98–1.16 |
| 5 | 1.13 | 0.97–1.28 |
| ≥6 or more | 1.31 | 1.10–1.57 |
| **Number of children <16 years old in the household** | | |
| 0 | 0.82 | 0.77–0.88 |
| 1 | 1.29 | 1.18–1.41 |
| 2 | 1.05 | 0.94–1.17 |
| 3 | 0.97 | 0.79–1.20 |
| 4 | 1.35 | 0.88–2.08 |
| ≥5 | 0.77 | 0.34–1.71 |
| **Number of members >65 years old in the household** | | |
| 0 | 0.84 | 0.76–0.92 |
| 1 | 1.33 | 1.18–1.50 |
| 2 | 0.97 | 0.84–1.14 |
| 3 | 0.99 | 0.44–2.24 |
| ≥4 | 1.82 | 0.54–6.03 |
| **Country of citizenship** | | |
| National of 13 sample countries | 0.78 | 0.68–0.91 |
| National of other EU[b] member state | 1.28 | 1.09–1.50 |
| National of non-EU country | 1.25 | 0.90–1.73 |
| **Country of birth** | | |
| Native of 13 sample countries | 1.02 | 0.89–1.17 |
| Native of other EU member state | 0.80 | 0.67–0.95 |
| Native of non-EU country | 1.31 | 1.06–1.61 |
| **Type of locality** | | |
| Densely populated area (cities and large towns) | 1.17 | 1.09–1.25 |
| Intermediate area (towns) | 0.92 | 0.86–0.99 |
| Thinly populated area (villages and rural) | 0.90 | 0.83–0.97 |
| **Completed level of education** | | |
| Primary or lower secondary education | 0.87 | 0.80–0.95 |



XSL•FO
RenderX

| | OR[a] | 95% CI |
|---|---|---|
| Upper secondary education | 1.08 | 1.01–1.16 |
| Tertiary education | 1.01 | 0.94–1.08 |
| **Labor status** | | |
| Employed or self-employed | 1.07 | 0.99–1.15 |
| Unemployed | 0.98 | 0.87–1.10 |
| Student | 0.96 | 0.87–1.05 |
| Not in the labor force (retired, inactive) | 0.94 | 0.86–1.03 |
| **Net monthly income range, (€)** | | |
| 1–1000 | 1.66 | 1.48–1.87 |
| 1001–2000 | 0.78 | 0.69–0.98 |
| 2001–3000 | 0.78 | 0.68–0.91 |
| 3001–4000 | 0.80 | 0.64–0.99 |
| ≥4001 | 0.85 | 0.66–1.12 |

[a]OR: odds ratio.

[b]EU: European Union.

From the viewpoint of residence and nationality, residence in other European Union countries (OR 1.28, 95% CI 1.09–1.50), and residence (OR 1.25, 95% CI 0.90–1.73) or birth (OR 1.31, 95% CI 1.06–1.61) outside the European Union determined higher probabilities of intensive eHealth usage. In contrast, European Internet users had a lower propensity toward more intensive eHealth usage if they had citizenship (OR 0.78, 95% CI 0.68–0.91) or were born in 1 of the 13 countries in the sample (OR 1.02, 95% CI 0.89–1.17). By municipality type, eHealth usage was more intensive among users residing in densely populated areas, such as cities and large towns (OR 1.17, 95% CI 1.09–1.25). Internet users residing in intermediate areas, such as towns (OR 0.92, 95% CI 0.86–0.99), or in less densely populated areas, such as village and rural areas (OR 0.90, 95% CI 0.83–0.97), had a lower propensity toward intensive eHealth usage.

Finally, European Internet users' educational levels and occupational category presented an inverted U shape in relation to more intensive eHealth usage. Regarding levels of completed education, the propensity toward intensive eHealth usage was greater among those with a secondary education (OR 1.08, 95% CI 1.01–1.16). In contrast, users with primary (OR 0.87, 95% CI 0.80–0.95) and tertiary (OR 1.01, 95% CI 0.94–1.08) education had a lower propensity. In terms of occupational category, the propensity toward intensive eHealth usage was greater among the employed or self-employed (OR 1.07, 95% CI 0.99–1.15). Users who were unemployed (OR 0.98, 95% CI 0.87–1.10), students (OR 0.96, 95% CI 0.87–1.05), or not in the labor force (OR 0.94, 95% CI 0.86–1.03) had lower probabilities of more intensive eHealth usage. In explaining more intensive eHealth usage as a consequence of users' net monthly income, the results suggest a higher propensity among the minimum wage stratum, earning ≤€1000 per month (OR 1.66, 95% CI 1.48–1.87).

## Discussion

The widespread use of ICTs in general and of the Internet in particular, together with the economic and social changes arising therefrom, are creating a fast-paced and significant change in relationships formed among the stakeholders of the health care system. One of the main manifestations of this disruptive process of change is the watering down of the traditional doctor-patient relationship model. Health Internet (eHealth) usage creates new dynamics that put the patient at the heart of the health care process. Doctor-patient interaction is no longer limited to time and place or to a few minutes in a doctor's office; nowadays, digital flows of information, communication, and knowledge go beyond the scope of health care centers and pervade the daily lives of citizens.

In this new context, the importance of evaluating the extent to which eHealth usage empowers citizens and involves them in their health status has been noted in the literature [30,39]. While there is considerable evidence in the literature about the predictors of some particular uses of eHealth, generally for population samples [41], attention has recently been drawn to the need to use more advanced methods and models to evaluate the participation of patients and citizens in the shared health care model that eHealth proposes [36].

This is why the goal of our study was to design and evaluate a predictive multidimensional model of eHealth usage, comprising 9 dimensions and 88 indicators. To that end, we used a broad sample of 13,000 European Internet users. Although we did not use a population sample, the results obtained are very useful, for two reasons. First, obtaining new evidence centered solely on Internet users allowed us to focus the analysis better, particularly with regard to inequalities (health status, sex, age, nationality, territory, education, and occupational category) that determine intensive eHealth usage. Second, the predictors we



XSL•FO
RenderX

obtained provided evidence that complements studies that have taken a population approach.

### eHealth Usage Composite Indicator

In recent years, eHealth usage has increased considerably [27,51]. More than half of the European population uses the Internet to look for health information [52], and more and more people are using it to access and manage their own personal health records [53], to buy health products and services, to communicate with their physicians [8,54], and to create digital content. In our study, we constructed a composite indicator using a 2-stage SEM methodology, and the results obtained are consistent with this evidence: they showed that, in 2011, 52.39% (6811/13,000) of European Internet users' eHealth usage was intensive (higher than the mean). The dimensions with more explanatory power in the eHealth usage composite indicator were health Internet attitudes, information health Internet usage, empowerment of health Internet users, and usefulness of health Internet usage.

### Health Status-Related Predictors of eHealth

Regarding eHealth predictors, while differences between European Internet users' perceived general health status and more intensive eHealth usage were not significant, long-term health problems or illnesses in the user or a family member did determine predictive power. European Internet users with long-term health problems or illnesses or receiving long-term treatment, or who had family members or cared for people with long-term health problems or illnesses had a greater propensity toward more intensive eHealth usage. Likewise, the study also highlighted that the existence of certain illnesses among the European Internet user population had high explanatory power with respect to intensive eHealth usage. These health problems or illnesses were diabetes, stroke or cerebral hemorrhage, cancer, and cataract. In contrast, users with health problems or illnesses related to chronic bronchitis and emphysema, and to osteoporosis had a lower propensity toward intensive eHealth usage.

These results, which are clearly consistent with other studies of social networking sites, virtual communities, and support group usage by patients with chronic illnesses [55], point to these patients' need for information and communication flows via eHealth to manage their health problems. The link between eHealth and chronic health problems determines the choice to develop specific practices in this field, and especially to provide those in this segment of the population (the chronically ill and caregivers) who are still not Internet users with greater digital competencies.

### Sociodemographic-Related Predictors of eHealth

Our results suggest that women, those aged 25–54 years, and households with more members, more children <16 years of age, and more members >65 years of age were most likely to use eHealth intensively. In contrast, men, people in the age groups 16–24 years and 54–74 years, and households with fewer members or with fewer dependents were less likely to use eHealth intensively.

The decisive importance of women [44], the middle age segments, and care of dependents is explained by the nature of health care in households and by the progressive aging of the population. It is important to underscore that women's role as health caregivers in the household clearly determines the usefulness of eHealth practices. In this respect, practices for fostering eHealth usage should consider the sex dimension more carefully. To a large extent, household eHealth usage arises through the health care role that families assign to women.

Aging of the population poses a broad set of challenges for health care systems, which a more widespread implementation of eHealth could help to meet. Without doubt, the main challenge for sustainable health that Europe faces over the coming years is the aging of the population. This is a complex mix of genetic, environmental, lifestyle, and socioeconomic factors, with the rates of associated chronic illnesses. Indeed, the European population is changing dramatically because of longer life expectancy and lower fertility rates. The number of European citizens over the age of 80 years is expected to double by 2025, which will give rise to increasingly complex needs in terms of clinical care, health care, and social care. In this context, eHealth practices could become one of the main tools for delivering health care to older citizens, especially through female caregivers. While the new patient-centered model has increasingly underscored the empowerment of patients and users in health care, the aging care model should be characterized by interaction between an active and informed patient or caregiver and a proactive and versatile medical team [56,57]. To that end, and given that the results obtained from this study show that middle-aged Internet users had a high propensity toward eHealth usage, it is essential to provide older caregivers who are still not Internet users with greater digital competencies.

From the perspective of nationality and territory, significant results were also obtained from the study. European Internet users had a greater propensity toward more intensive eHealth usage if they resided in other European Union countries or outside the European Union, and if they were born outside the European Union. Similarly, European Internet users' residence in densely populated areas (cities or large towns) also better predicted eHealth usage. In this context, a fairer promotion of eHealth usage in Europe should also consider the territorial dimension, with special emphasis on connecting national health systems and a greater Internet presence and usage in less densely populated areas.

Finally, the results obtained also provide us with significant information about educational, occupational, and income categories, which are crucial for redressing some of the social inequalities in eHealth usage. Users' educational levels explain more intensive eHealth usage, in an inverted U form. Thus, users with a secondary education had a greater propensity toward intensive eHealth usage. In this sense, the study provides new evidence (beyond population studies) in relation to middle-educated (secondary education) Internet users, who perceived the usefulness of eHealth usage. The education dimension also determines a new area of health inequality, and hence the need to promote Internet usage among the less educated population. The results related to occupational and income categories suggest a higher propensity among the



XSL•FO
RenderX

employed or self-employed and among the minimum wage stratum earning ≤€1000 per month. Users who were integrated into the labor market, whether self-employed or employed, clearly had a greater propensity, whereas those who were not (students, unemployed, and not in the labor force) had a lower propensity to use eHealth. In this context, in order to achieve a more equitable eHealth usage, Internet usage among groups not actively integrated into the labor market should be promoted more vigorously. Regarding income, and in order to overcome inequalities, promoting eHealth usage skills (especially through education and learning) for workers with lower wages would also be very useful.

### Limitations

Our study has several limitations. First, there was a time lag between the year we obtained the data and the year we wrote the paper. However, we felt that the availability of a single database of 13,000 Internet users in Europe deserved an analysis despite the time lag. In future research, and as they become available, we will use newer data and introduce dynamic comparisons. Second, the study provides information only from the perspective of health users. In the future, we intend to address the issue of eHealth usage by health professionals. By doing so, we will be able to improve our multidimensional approach and obtain results and conclusions for all actors involved in eHealth usage. Third, the empirical methodology could also be improved by looking at the intensity of eHealth usage (not simply usage or mean usage) and at a higher number of predictors.

### Conclusions

The results obtained highlight the need for more in-depth research to be conducted into the link between eHealth usage and predictors, and the different health care systems in Europe. By doing so, it will be possible to increase the resolution of our results and to establish whether the intensity of eHealth usage varies depending on the health care systems, or the extent to which health care systems determine the prediction of eHealth usage. Similarly, strategic and public policy actions resulting from the research could be adapted more precisely to each health care system. Finally, the study results could be supplemented by the construction of a composite indicator of eHealth usage by health care professionals. The design, validation, and prediction of composite indicators of eHealth usage that take into consideration the perspectives of both users (ie, patients) and professionals in the different European health care systems would provide us with a very comprehensive view of the issue and would allow us to round off our multidimensional approach. We shall focus our efforts on all of these approaches in the near future.

### Acknowledgments

This work used data from the Strategic Intelligence Monitor on Personal Health Systems Phase 2 (SIMPHS2) research project, carried out by the Institute for Prospective Technological Studies (IPTS) in cooperation with the European Commission Directorate General for Information Society and Media. The funding sources had no involvement in this study. The authors would like to thank Ioannis Maghiros, Head of Unit, Information Society, at IPTS for his support. Open access microdata are available from the Joint Research Centre (http://is.jrc.ec.europa.eu/pages/TFS/SIMPHS2deliverables.html).

### Authors' Contributions

All authors contributed substantially to the design, data analysis, and interpretation of the findings. Joan Torrent-Sellens participated in formulating the research question, study design, literature review, data analysis and statistical modelling, interpretation of the findings, and drafting the manuscript. Ángel Díaz-Chao participated in data analysis and statistical modelling. Ivan Soler-Ramos participated in the design and data analysis. Francesc Saigí-Rubió contributed to formulating the research question, study design, literature review, interpretation of results, and drafting the manuscript. He is the guarantor of the paper. All the authors have read, revised, and approved the final manuscript.

### Conflicts of Interest

None declared.

### Multimedia Appendix 1

Statistical information based on SIMPHS2 online survey.

[PDF File (Adobe PDF File), 25KB - jmir_v18i7e188_app1.pdf ]

### Multimedia Appendix 2

Strategic Intelligence Mapping on Personal Health Systems Phase 2 (SIMPHS2) Questionnaire.

[PDF File (Adobe PDF File), 145KB - jmir_v18i7e188_app2.pdf ]

### Multimedia Appendix 3

Health Internet uses descriptive statistics and frequency statistics. 2011.



XSL•FO
RenderX

[PDF File (Adobe PDF File), 37KB - jmir_v18i7e188_app3.pdf ]

### Multimedia Appendix 4

Health care Internet uses descriptive statistics and frequency statistics. 2011.

[PDF File (Adobe PDF File), 35KB - jmir_v18i7e188_app4.pdf ]

### Multimedia Appendix 5

Drivers of health care Internet uses and frequency statistics. 2011.

[PDF File (Adobe PDF File), 34KB - jmir_v18i7e188_app5.pdf ]

### Multimedia Appendix 6

Barriers of health care Internet uses descriptive statistics and frequency statistics. 2011.

[PDF File (Adobe PDF File), 35KB - jmir_v18i7e188_app6.pdf ]

### Multimedia Appendix 7

Usefulness of health Internet uses descriptive statistics and frequency statistics. 2011.

[PDF File (Adobe PDF File), 37KB - jmir_v18i7e188_app7.pdf ]

### Multimedia Appendix 8

ICT uses descriptive statistics and frequency statistics. 2011.

[PDF File (Adobe PDF File), 37KB - jmir_v18i7e188_app8.pdf ]

### Multimedia Appendix 9

Information health Internet uses descriptive statistics and frequency statistics. 2011.

[PDF File (Adobe PDF File), 35KB - jmir_v18i7e188_app9.pdf ]

### Multimedia Appendix 10

Health Internet attitudes descriptive statistics and frequency statistics. 2011.

[PDF File (Adobe PDF File), 35KB - jmir_v18i7e188_app10.pdf ]

### Multimedia Appendix 11

Empowerment of health Internet user's descriptive statistics and frequency statistics. 2011.

[PDF File (Adobe PDF File), 34KB - jmir_v18i7e188_app11.pdf ]

## References

1. Lu Y, Xiao Y, Sears A, Jacko JA. A review and a framework of handheld computer adoption in healthcare. Int J Med Inform 2005 Jun;74(5):409-422. [doi: 10.1016/j.ijmedinf.2005.03.001] [Medline: 15893264]
2. European Commission. eHealth for Safety Report. Luxembourg: Office for Official Publications of the European Communities; 2007.
3. Eysenbach G. What is e-health? J Med Internet Res 2001 Jun;3(2):E20. [doi: 10.2196/jmir.3.2.e20] [Medline: 11720962]
4. Department of Health. Delivering 21st Century IT Support for the NHS. London: DOH; 2002:1-31.
5. Oh H, Rizo C, Enkin M, Jadad A. What is eHealth (3): a systematic review of published definitions. J Med Internet Res 2005;7(1):e1 [FREE Full text] [doi: 10.2196/jmir.7.1.e1] [Medline: 15829471]
6. 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]
7. Dansky K, Thompson D, Sanner T. A framework for evaluating eHealth research. Eval Program Plann 2006 Nov;29(4):397-404. [doi: 10.1016/j.evalprogplan.2006.08.009] [Medline: 17950868]



XSL•FO
RenderX

8. Fox S, Duggan M. Health Online 2013. Washington, DC: Pew Internet & American Life Project; 2013 Jan 15. URL: http://www.pewinternet.org/files/old-media//Files/Reports/PIP_HealthOnline.pdfpx [accessed 2016-07-06] [WebCite Cache ID 6ingAGNFu]
9. Campbell R. Consumer health, patient education, and the Internet. Internet Journal of Health 2001;2(2):001-008.
10. Eysenbach G. Medicine 2.0: social networking, collaboration, participation, apomediation, and openness. J Med Internet Res 2008;10(3):e22 [FREE Full text] [doi: 10.2196/jmir.1030] [Medline: 18725354]
11. Kreps G, Neuhauser L. New directions in eHealth communication: opportunities and challenges. Patient Educ Couns 2010;78(3):329-336.
12. Lee Y, Lin J. Do patient autonomy preferences matter? Linking patient-centered care to patient-physician relationships and health outcomes. Soc Sci Med 2010 Nov 01;71(10):1811-1818. [doi: 10.1016/j.socscimed.2010.08.008] [Medline: 20933316]
13. Veroff D, Marr A, Wennberg D. Enhanced support for shared decision making reduced costs of care for patients with preference-sensitive conditions. Health Aff (Millwood) 2013;32(2):285-293.
14. Bertakis K, Azari R. Patient-centered care is associated with decreased health care utilization. J Am Board Fam Med 2011;24(3):229-239.
15. Greaves F, Pape U, King D, Darzi A, Majeed A, Wachter R, et al. Associations between Internet-based patient ratings and conventional surveys of patient experience in the English NHS: an observational study. BMJ Qual Saf 2012 Jul;21(7):600-605. [doi: 10.1136/bmjqs-2012-000906] [Medline: 22523318]
16. Hu X, Bell RA, Kravitz RL, Orrange S. The prepared patient: information seeking of online support group members before their medical appointments. J Health Commun 2012;17(8):960-978. [doi: 10.1080/10810730.2011.650828] [Medline: 22574697]
17. Gao G, McCullough J, Agarwal R, Jha A. A changing landscape of physician quality reporting: analysis of patients' online ratings of their physicians over a 5-year period. J Med Internet Res 2012;14(1):e38 [FREE Full text] [doi: 10.2196/jmir.2003] [Medline: 22366336]
18. Kelders SM, van Gemert-Pijnen JE, Werkman A, Seydel ER. Evaluation of a web-based lifestyle coach designed to maintain a healthy bodyweight. J Telemed Telecare 2010;16(1):3-7. [doi: 10.1258/jtt.2009.001003] [Medline: 20086259]
19. Verhoeven F, Tanja-Dijkstra K, Nijland N, Eysenbach G, van Gemert-Pijnen JE. Asynchronous and synchronous teleconsultation for diabetes care: a systematic literature review. J Diabetes Sci Technol 2010 May;4(3):666-684 [FREE Full text] [Medline: 20513335]
20. Commission to the European Parliament, The Council, The European Economic and Social Committee, The Committee of the Regions. eHealth Action Plan 2012-2020: Innovative Healthcare for the 21st Century. Brussels, Belgium: European Commission; 2012. URL: http://ec.europa.eu/health/ehealth/docs/com_2012_736_en.pdf [accessed 2016-06-30] [WebCite Cache ID 6ikm7xJ0l]
21. Politi MC, Wolin KY, Légaré F. Implementing clinical practice guidelines about health promotion and disease prevention through shared decision making. J Gen Intern Med 2013 Jun;28(6):838-844. [Medline: 23307397]
22. Grol R, Grimshaw J. From best evidence to best practice: effective implementation of change in patients' care. Lancet 2003 Oct 11;362(9391):1225-1230. [doi: 10.1016/S0140-6736(03)14546-1] [Medline: 14568747]
23. Nijland N, van Gemert-Pijnen JE, Boer H, Steehouder M, Seydel E. Evaluation of internet-based technology for supporting self-care: problems encountered by patients and caregivers when using self-care applications. J Med Internet Res 2008;10(2):e13 [FREE Full text] [doi: 10.2196/jmir.957] [Medline: 18487137]
24. Kelders S, Van Gemert-Pijnen JE, Werkman A, Nijland N, Seydel E. Effectiveness of a Web-based intervention aimed at healthy dietary and physical activity behavior: a randomized controlled trial about users and usage. J Med Internet Res 2011;13(2):e32 [FREE Full text] [doi: 10.2196/jmir.1624] [Medline: 21493191]
25. Xie B, Wang M, Feldman R. Preferences for health information and decision-making: development of the Health Information Wants (HIW) Questionnaire. In: Proceedings of the 2011 iConference. New York, NY: ACM; 2011 Presented at: iConference; February 8-11, 2011; Seattle, Washington.
26. Lustria ML, Smith S, Hinnant C. Exploring digital divides: an examination of eHealth technology use in health information seeking, communication and personal health information management in the USA. Health Inform J 2011 Sep;17(3):224-243. [doi: 10.1177/1460458211414843] [Medline: 21937464]
27. Beckjord E, Finney Rutten LJ, Squiers L, Arora N, Volckmann L, Moser R, et al. Use of the internet to communicate with health care providers in the United States: estimates from the 2003 and 2005 Health Information National Trends Surveys (HINTS). J Med Internet Res 2007;9(3):e20 [FREE Full text] [doi: 10.2196/jmir.9.3.e20] [Medline: 17627929]
28. Kontos EZ, Emmons KM, Puleo E, Viswanath K. Contribution of communication inequalities to disparities in human papillomavirus vaccine awareness and knowledge. Am J Public Health 2012 Oct;102(10):1911-1920. [doi: 10.2105/AJPH.2011.300435] [Medline: 22970692]
29. Thackeray R, Crookston BT, West JH. Correlates of health-related social media use among adults. J Med Internet Res 2013;15(1):e21 [FREE Full text] [doi: 10.2196/jmir.2297] [Medline: 23367505]
30. Gibbons MC, Fleisher L, Slamon RE, Bass S, Kandadai V, Beck JR. Exploring the potential of Web 2.0 to address health disparities. J Health Commun 2011;16 Suppl 1:77-89. [doi: 10.1080/10810730.2011.596916] [Medline: 21843097]




XSL•FO
RenderX

31. Zach L, Dalrymple PW, Rogers ML, Williver-Farr H. Assessing internet access and use in a medically underserved population: implications for providing enhanced health information services. Health Info Libr J 2012 Mar;29(1):61-71 [FREE Full text] [doi: 10.1111/j.1471-1842.2011.00971.x] [Medline: 22335290]
32. Eysenbach G, Köhler C. How do consumers search for and appraise health information on the world wide web? Qualitative study using focus groups, usability tests, and in-depth interviews. BMJ 2002 Mar 9;324(7337):573-577 [FREE Full text] [doi: 10.2196/jmir.5.4.e33] [Medline: 11884321]
33. Van Der Meijden MJ, Tange HJ, Troost J, Hasman A. Determinants of success of inpatient clinical information systems: a literature review. J Am Med Inform Assoc 2003;10(3):235-243. [doi: 10.1197/jamia.M1094] [Medline: 12626373]
34. Kern J. Evaluation of teleconsultation systems. Int J Med Inform 2006;75(3-4):330-334. [doi: 10.1016/j.ijmedinf.2005.08.001] [Medline: 16143564]
35. Wootton R, Hebert M. What constitutes success in telehealth? J Telemed Telecare 2001;7 Suppl 2:3-7. [Medline: 11747643]
36. James S. The Content of Social Explanation. Cambridge, UK: Cambridge University Press; 1984.
37. Honderich T. The Oxford Companion to Philosophy: New Edition. New York, NY: Oxford University Press; 2005.
38. Lawrence C, Weisz G. Greater Than the Parts: Holism in Biomedicine, 1920-1950. New York, NY: Oxford University Press; 1998.
39. Kontos E, Emmons K, Puleo E, Viswanath K. Communication inequalities and public health implications of adult social networking site use in the United States. J Health Commun 2010;15 Suppl 3:216-235 [FREE Full text] [doi: 10.1080/10810730.2010.522689] [Medline: 21154095]
40. Chou WS, 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]
41. Kontos E, Blake KD, Chou WS, Prestin A. Predictors of eHealth usage: insights on the digital divide from the Health Information National Trends Survey 2012. J Med Internet Res 2014;16(7):e172 [FREE Full text] [doi: 10.2196/jmir.3117] [Medline: 25048379]
42. Hargittai E. Digital na(t)ives? Variation in Internet skills and uses among members of the “net generation”. Sociol Inquiry 2010;80(1):92-113.
43. Kontos EZ, Bennett GG, Viswanath K. Barriers and facilitators to home computer and internet use among urban novice computer users of low socioeconomic position. J Med Internet Res 2007;9(4):e31 [FREE Full text] [doi: 10.2196/jmir.9.4.e31] [Medline: 17951215]
44. Wang J, Bennett K, Probst J. Subdividing the digital divide: differences in internet access and use among rural residents with medical limitations. J Med Internet Res 2011;13(1):e25 [FREE Full text] [doi: 10.2196/jmir.1534] [Medline: 21371989]
45. Lupiáñez F, Maghiros I, Abadie F. Citizens and ICT for health in 14 European countries: results from an online panel. Final report of the “Strategic Intelligence Monitor on Personal Health Systems Phase 2 (SIMPHS2)”. European Commission JRC Scientific and Policy Reports URL: http://ftp.jrc.es/EURdoc/JRC71142.pdf [accessed 2016-07-04] [WebCite Cache ID 6iklRTgVU]
46. Eysenbach G. Improving the quality of Web surveys: the Checklist for Reporting Results of Internet E-Surveys (CHERRIES). J Med Internet Res 2004 Sep 29;6(3):e34 [FREE Full text] [doi: 10.2196/jmir.6.3.e34] [Medline: 15471760]
47. Farber S. Factors Influencing eHealth Adoption by Dutch Hospitals: An Empirical Study [doctoral dissertation]. Delft University of Technology: Delft; 2014 Oct 10. URL: http://repository.tudelft.nl/islandora/search/collection%253Air?collection=research [accessed 2016-07-06] [WebCite Cache ID 6inlU4yvp]
48. Cho J, Park D, Lee HE. Cognitive factors of using health apps: systematic analysis of relationships among health consciousness, health information orientation, eHealth literacy, and health app use efficacy. J Med Internet Res 2014;16(5):e125 [FREE Full text] [doi: 10.2196/jmir.3283] [Medline: 24824062]
49. Marcoulides G, Moustaki I. Latent Variable and Latent Structure Models. New York, NY: Routledge; 2012.
50. Pagliari C, Sloan D, Gregor P, Sullivan F, Detmer D, Kahan J, et al. What is eHealth (4): a scoping exercise to map the field. J Med Internet Res 2005;7(1):e9 [FREE Full text] [doi: 10.2196/jmir.7.1.e9] [Medline: 15829481]
51. Eysenbach G, Kohler C. What is the prevalence of health-related searches on the World Wide Web? Qualitative and quantitative analysis of search engine queries on the internet. In: AMIA Annual Symposium Proceedings. 2003 Presented at: AMIA Annual Symposium; November 8-12, 2003; Washington, DC p. 225-229 URL: http://europepmc.org/abstract/MED/14728167
52. Kummervold P, Chronaki C, Lausen B, Prokosch H, Rasmussen J, Santana S, et al. eHealth trends in Europe 2005-2007: a population-based survey. J Med Internet Res 2008;10(4):e42 [FREE Full text] [doi: 10.2196/jmir.1023] [Medline: 19017584]
53. Ball M, Smith C, Bakalar R. Personal health records: empowering consumers. J Healthc Inf Manag 2007;21(1):76-86. [Medline: 17299929]
54. Adler K. Web portals in primary care: an evaluation of patient readiness and willingness to pay for online services. J Med Internet Res 2006;8(4):e26. [doi: 10.2196/jmir.8.4.e26]
55. West D. Enabling Personalized Medicine through Health Information Technology: Advancing the Integration of Information. Washington, DC: Brookings Institution; 2011 Jan 28. URL: http://www.brookings.edu/research/papers/2011/01/28-personalized-medicine-west [accessed 2016-07-01] [WebCite Cache ID 6iklx9knq]




XSL•FO
RenderX

56. Glasgow RE, Orleans CT, Wagner EH. Does the chronic care model serve also as a template for improving prevention? Milbank Q 2001;79(4):579-612, iv [FREE Full text] [Medline: 11789118]
57. Wagner EH, Bennett SM, Austin BT, Greene SM, Schaefer JK, Vonkorff M. Finding common ground: patient-centeredness and evidence-based chronic illness care. J Altern Complement Med 2005;11 Suppl 1:S7-15. [doi: 10.1089/acm.2005.11.s-7] [Medline: 16332190]

## Abbreviations

**CFI:** comparative fit index
**ICT:** information and communication technology
**IFI:** incremental fit index
**NFI:** normed fit index
**OR:** odds ratio
**RFI:** relative fit index
**RMSEA:** root mean square error of approximation
**SEM:** structural equation modelling
**SIMPHS2:** Strategic Intelligence Monitor on Personal Health Systems Phase 2
**TLI:** Tucker-Lewis index

*Edited by G Eysenbach; submitted 11.02.16; peer-reviewed by JY Kim, M Johnston; comments to author 03.04.16; revised version received 11.05.16; accepted 30.05.16; published 22.07.16.*

*Please cite as:*
*Torrent-Sellens J, Díaz-Chao Á, Soler-Ramos I, Saigí-Rubió F*
*Modelling and Predicting eHealth Usage in Europe: A Multidimensional Approach From an Online Survey of 13,000 European Union Internet Users*
*J Med Internet Res 2016;18(7):e188*
*URL: http://www.jmir.org/2016/7/e188/*
*doi:10.2196/jmir.5605*
*PMID:27450189*

©Joan Torrent-Sellens, Ángel Díaz-Chao, Ivan Soler-Ramos, Francesc Saigí-Rubió. Originally published in the Journal of Medical Internet Research (http://www.jmir.org), 22.07.2016. This is an open-access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/2.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on http://www.jmir.org/, as well as this copyright and license information must be included.



XSL•FO
RenderX

Review

# The Personal Emergency Response System as a Technology Innovation in Primary Health Care Services: An Integrative Review

Randi Stokke[1,2], RN, MEd

[1]Centre for Care Research, Norwegian University of Science and Technology (NTNU), Gjøvik, Norway

[2]The Centre for Innovation in Services, Lillehammer University College, Lillehammer, Norway

**Corresponding Author:**
Randi Stokke, RN, MEd
Centre for Care Research
Norwegian University of Science and Technology (NTNU)
Postbox 191
Gjøvik, 2802
Norway
Phone: 47 95158088
Fax: 47 61135170
Email: randi.stokke@ntnu.no

## Abstract

**Background:** Most western countries are experiencing greater pressure on community care services due to increased life expectancy and changes in policy toward prioritizing independent living. This has led to a demand for change and innovation in caring practices with an expected increased use of technology. Despite numerous attempts, it has proven surprisingly difficult to implement and adopt technological innovations. The main established technological innovation in home care services for older people is the personal emergency response system (PERS), which is widely adopted and used throughout most western countries aiming to support "aging safely in place."

**Objective:** This integrative review examines how research literature describes use of the PERS focusing on the users' perspective, thus exploring how different actors experience the technology in use and how it affects the complex interactions between multiple actors in caring practices.

**Methods:** The review presents an overview of the body of research on this well-established telecare solution, indicating what is important for different actors in regard to accepting and using this technology in community care services. An integrative review, recognized by a systematic search in major databases followed by a review process, was conducted.

**Results:** The search resulted in 33 included studies describing different actors' experiences with the PERS in use. The overall focus was on the end users' experiences and the consequences of having and using the alarm, and how the technology changes caring practices and interactions between the actors.

**Conclusions:** The PERS contributes to safety and independent living for users of the alarm, but there are also unforeseen consequences and possible improvements in the device and the integrated service. This rather simple and well-established telecare technology in use interacts with the actors involved, creating changes in daily living and even affecting their identities. This review argues for an approach to telecare in which the complexity of practice is accounted for and shows how the plug-and-play expectations producers tend to generate is a simplification of the reality. This calls for a recognition that place and actors matter, as does a sensitivity to technology as an integrated part of complex caring practices.

***(J Med Internet Res 2016;18(7):e187)*** doi:10.2196/jmir.5727

**KEYWORDS**

home care services; caring practice; personal emergency alarm system; PERS; safety alarm; social alarm; telecare; review



XSL•FO
RenderX

## Introduction

### Background

Western societies have an ageing population due to increased life expectancy and large cohorts in the postwar years, presenting growing challenges to long-term care services [1,2]. Independent living for older people is a policy priority in western countries [3], and this includes active ageing and maintenance of quality of life [4,5].

Meeting such demands through technology innovation is one suggested solution. Technological innovations in community care services are highly regarded, even though integration and adoption has proven difficult and many projects never pass the pilot stage [6,7]. Studying how different actors perceive and experience existing technologies in use is one way of providing a richer and more nuanced view of what promotes or inhibits adoption of new technology innovations [8]. This review seeks to do so by exploring research on the personal emergency response system (PERS). The PERS is a widespread, integrated, and accepted technology innovation in care practices. Through focusing on the users` perspectives, this review explores how different actors experience the technology in use and how it affects the complex interactions between multiple actors in caring practices.

### The Personal Emergency Response System

The PERS has proved sustainable over time. Since its launch in the late 1970s, it is widely used and spread throughout most western societies [9]. For example, there are approximately 1.4 million users in the United Kingdom [10] and 74,000 users in Norway [11]. Even after many years of use, no review article summarizing use of the PERS research has been identified. In fact, the research literature on the subject is rather sparse [12].

The PERS is a technological device and an integrated service, embodying three generations of alarm devices as a result of technical development, although some characteristics remain unchanged. The first generation alarm device had a unit placed centrally in the home, with a switch or a pull cord to use in emergency. The second generation has in addition a pendant, a necklace/wristband with a button that the user can press in case of emergency. This allows open communication between the user and a responder through the main unit, enabling the responders to effect a proper response. The range of the pendant is normally inside the home and partly outside. The third and newest generation of the PERS has the potential to incorporate a range of devices (eg, automatic fall alarms, fire alarms, and blood pressure devices [10,13]), providing remote care [14].

It is mainly the second and third generation versions that are in use today, although implementation of additional alarms, devices, and services has proven difficult and is done on a much smaller scale than expected with a slowly growing market. Thus, the PERS might be described as a foundation for further safety and monitoring telecare.

The organization of the PERS as a service varies from private arrangements, where the alarm goes to a nominated contact, to small or large public or private call centers answering and effecting proper responses [1,10]. The PERS as a service system is complex, dealing with a variety of contexts and services [15], and its organization and use are part of integrated caring practices with multiple actors. The different actors are, among others, the end user and their relatives and neighbors, home care nurses, and telecare facilitators. They are all users of the PERS but have different experiences, roles, meanings, and relationships with each other and the technology. The end user of the PERS with the alarm in his/her home is usually an older person [11,16].

### Use of Technology in Caring Practices

Policy makers and advocates of such alarm systems often describe telecare technologies such as the PERS as ‘‘plug-n-play’’ solutions with placement of devices at home providing help in an effective way, enhancing quality of life, and reducing costs for the care service [15]. However, the use of telecare tends to be more complex than such promises suggest [17]. There are many indications that moving away from the rather naïve technological determinism, where telecare technologies are simply viewed as plug-and-play devices, and instead acknowledging the complexity in technology practices would provide a more accurate view of practice.

### Theorizing Use of Technology in Caring Practices

When studying practice, Nicolini [18] argues that the purpose of social science is to open up for a rich and nuanced understanding of practice [18]. He argues that there is no such thing as a unified practice theory and suggests using what he calls a toolkit approach by mobilizing different aspects of similar theories when exploring practice. This enables enriched understanding of what is going on. He suggests what he calls a “theory-method package,” which when utilized in this review involves zooming in on the practice of the PERS in use as displayed in the included articles, and zooming out following trails of connections. By zooming out, it is possible to draw on the local practice of the PERS in use to acquire a wider picture of technology in use in caring practices. Scholars such as Nelly Oudshoorn [19-21], Jeanette Pols [15,22,23], and Davide Nicolini [18,24,25] have studied different kinds of telecare that will provide tools for zooming out, exploring and theorizing technology in use in caring practices. Both Nicolini and Oudshoorn are inspired by science and technology studies (STS) and the fields of human geographies aiming to bridge these approaches by focusing on how place is important when shaping user and technology relations [18,20,24]. Technology, actors, society, and place must be thought of together since they are coconstructed, and technology is by definition technology in use.

Three main questions exploring how technology in use is integrated in caring practices arise from what research tells us about the users’ experiences with the PERS:

1. What has research focused on when studying use of the PERS over time?

2. How do actors in home care practices experience, integrate, and relate to the PERS in everyday life?

3. How does this established technology in use influence and affect caring practices?



XSL•FO
RenderX

## Methods

An integrative review was conducted, characterized by explicit, rigorous, and transparent methodology using a systematic search, but allowing the inclusion of research with diverse methodologies and a broader range of studies [26]. Ethical approval was not required as this was secondary research. Textbox 1 presents inclusion criteria applied to articles before the searches.

**Textbox 1.** Inclusion criteria for articles to be included in the review.

Inclusion criteria:

- Articles dealing with older people's attitudes, experiences, interactions, feelings, use and nonuse, consequences, and effects of use of a personal emergency response system (PERS) in home care
- Peer-reviewed articles from academic journals describing and focusing on different aspects of the PERS in use rather than articles with a main focus on further technology innovation
- Articles written in the English language, no limitation in publication period, and no methodological restrictions

A systematic search of relevant terms was conducted in relevant databases and search engines. The search strategies and results are presented in Figure 1. All articles were reviewed according to the inclusion criteria. A thorough description of the steps describing the research strategy process is described in Multimedia Appendix 1. A data extraction sheet was a useful tool for quality in assessing the articles.

A descriptive, integrative, thematic analysis as described by Whittemore and Knafl [26] was used to analyze the articles. This required ordering, coding, categorizing, and summarizing data [26]. Table 1 presents a comparative and systematic organization of the included studies [9,16,22-52].

The next steps were exploring and displaying the extracted data around different variables and subgroups looking for patterns, themes, and relationships, and then drawing a map of the essential identified themes. This was followed by abstracting and grouping themes into categories, aiming to subsume the particulars into more general findings. To ensure quality, the included articles were checked to verify for accuracy and conformability. Uncertainties throughout the process were discussed with a group of supervisors. Methodological considerations are described in Multimedia Appendix 2.



XSL•FO
RenderX

**Table 1.** An overview of the included articles in this integrative review.

| Article | Country | Methods | Main findings | Term used |
|---|---|---|---|---|
| Boström et al 2011 [42] | | | | |
| | Sweden | Focus group interviews with PERS[a] users | The participants' opinions and feelings with the PERS related to five themes: safety, anxiety, satisfaction, information, and older persons as active innovators. | PERS |
| De San Miguel and Lewin 2008 [43] | | | | |
| | Australia | Mail survey to 2610 PERS users | Clients reported impacts on emergency response, living independently, sense of security and anxiety, and when and where they wear their alarm. | Personal alarms |
| Fallis et al 2007 [33] | | | | |
| | Canada | Mixed-method design, survey, and qualitative feedback | Need for improvement. The PERS gave sense of security, comfort, and reassurance, with high satisfaction with service during an emergency. | PERS |
| Farquhar et al 1992 [47] | | | | |
| | Australia | Assessment intervention with 125 persons | Respondents described high satisfaction with the alarm. Total of 38% gave up the alarm; 62% never used the alarm, but 84% felt they required it. | Personal emergency alarms |
| Fisk 1995 [38] | | | | |
| | United Kingdom and Canada | Qualitative interviews with 38 users from Oldham and Ottawa | A majority experienced a feeling of security; 40-50% had used system in emergency. The alarm was not always used in emergencies. | Personal response services |
| Fleming and Brayne 2008 [48] | | | | |
| | United Kingdom | 1-year follow-up of 110 patients | Total of 54% of reported falls happened when person was alone; 80% did not use alarm to summon help. Users described different barriers. | Call alarm system |
| Heinbüchner et al 2010 [9] | | | | |
| | Germany | 333 PERS users approached; response rate 19.6% | Respondents were satisfied with their device, although 24% never wore the pendant. The PERS was not activated by 83% of the persons who fell. | PERS |
| Hyer and Rudick 1994 [44] | | | | |
| | United States | Telephone survey of 117 patients monitored; maximum 1 year | One-third of PERS users requested emergency assistance (60 calls); significant cost savings; high patient satisfaction. | PERS |
| Johnston et al 2010 [35] | | | | |
| | Australia | 31 semistructured interviews | Identified four subgroups: 1) used alarm effectively, 2) had alarm, but not used effectively, 3) no alarm, but were receptive, 4) no alarm and would not use it. | Personal alarms |
| Johnston et al 2010 [37] | | | | |
| | Australia | 1-month retrospective audit of 1700 cases (alarms) | Difficult to separate false alarm from emergencies. Personal alarm might be helpful for people living alone, when alarm is accepted, understood, and used effectively. | Personal alarms |
| Lee et al 2007 [45] | | | | |
| | Canada | RCT[b]; recruited after admitted to ED[c] after fall | There was no difference in mean change in anxiety between the groups. Alarm user had decreased fear of falling. | PERS |
| Levine and Tideiksaar 1995 [49] | | | | |
| | United States | Structured interviews; 106 participants | Total of 45% of respondents were fully compliant; identified factors that increased compliance. | PERS |



XSL•FO
RenderX

| Article | Country | Methods | Main findings | Term used |
|---|---|---|---|---|
| Mann et al 2005 [50] | | | | |
| | United States | Surveyed 606 people; users and nonusers of PERS | Total of 92.7% were satisfied with their PERS; 84.3% rated their PERS as very important. The major reason for potential use was falling and feeling ill. | PERS |
| McWhirter 1987 [51] | | | | |
| | United Kingdom | Quantitative client register questionnaire; 667 females, 194 males | Main reason for referral: problems with mobility (45.6%) and falls (43.4%); 40% of all calls were false alarms. | A dispersed alarm system |
| Melkas 2003 [55] | | | | |
| | Finland and Sweden | 40 interviews with service personnel | The study is mainly about information systems around use of the PERS; bottlenecks are identified. | Safety telephone services |
| Melkas 2010 [56] | | | | |
| | Finland | Human impact assessment methodologies | Total of 8 care workers at 8 workplaces. Changes, problems, and strengths related to information environment; improving information environment. | Safety telephone services |
| Nyman and Victor 2014 [41] | | | | |
| | United Kingdom | A secondary analysis from an English study of ageing | Investigated self-reported users of personal call alarms among 3091 adults aged 65+. From a large sample of those aged 65+, use of call alarm was rare. | Personal call alarms |
| Olsson et al 2012 [39] | | | | |
| | Sweden | Interview with 14 spouses of persons with dementia | Total of 4 spouses had safety alarm; used for different purposes (eg, if person with dementia had fallen or suddenly fell ill and they needed help). | Safety alarm |
| Pekkarinen and Melkas 2010 [16] | | | | |
| | Finland | Mixed methods; qualitative interviews; survey with users and personnel | Describing different “potholes” in the technology, service, process, organization, marketing, and ethics and how these can be dealt with. | Safety alarm systems |
| Porter 2003 [27] | | | | |
| | United States | 56 qualitative interviews with 8 widows | Experiences of having the PERS. The findings were a basis for considering the potential influences of having a PERS on older people’s well-being. | PERS |
| Porter 2008 [28] | | | | |
| | United States | Phenomenology; semistructured interviews with 14 women | How the PERS influenced what older people would do if an intruder got in their house. | PERS |
| Porter and Lasiter 2012 [29] | | | | |
| | United States | Phenomenology; part of a larger RHQ[d] project; 95 interviews with 25 women | Life-world being influenced by a peer’s situation regarding adopting or using a PERS for reaching help quickly. | PERS |
| Porter 2005 [30] | | | | |
| | United States | Phenomenology; interviews with 7 frail women during 3 years | The women’s experiences of PERS; a description of temporizing about the PERS button—deciding when to wear it and whether to use it. | PERS |
| Porter 2002 [31] | | | | |
| | United States | Phenomenology; part of a longitudinal study; 71 interviews of 11 widows | Experiences of not having the PERS; exploring reasons and barriers for PERS use. | PERS |



XSL•FO
RenderX

| Article | Country | Methods | Main findings | Term used |
|---|---|---|---|---|
| Porter et al 2013 [32] | | | | |
| | United States | Phenomenology; 99 interviews with 23 women | PERS subscribers' and nonsubscribers' intentions and context differ relative to reaching help quickly (RHQ). | PERS |
| Premik et al 1997 [53] | | | | |
| | Slovenia | Quantitative data from the PERS | Total of 18,500 alarm calls in 4 years; 2.1% health related. The alarm could be a basic communication device for older people. | Community social alarm system |
| Raappana et al 2007 [57] | | | | |
| | Finland | Human impact assessment methodology; 8 workplaces, 78 care workers | Safety alarms might be useful both for administration and actual care work. | Safety alarm system |
| Roush and Teasdale 2011 [34] | | | | |
| | United States and Canada | Survey; 267 older persons | PERS users utilized emergency departments twice as often as those without. Strong relation between access to a PERS, sense of security, and higher levels of well-being. | PERS |
| Roush et al 1995 [40] | | | | |
| | United States and Canada | Hospital utilization rates; 106 patients; 1-year follow-ups | PERS users had a significant decrease in per-person hospital admissions and inpatient days. No significant differences in ED visits. | PERS |
| Sjölinder et al 2014 [52] | | | | |
| | Sweden | Mixed-methods survey, interviews, and focus groups | The municipalities' knowledge about the new technology was deficient. Focuses on possibilities for using alarms outside. | Social alarm system |
| Tinker 1993 [36] | | | | |
| | United Kingdom | Literature summary from two major reports | Summary findings from two reports. | Dispersed alarms |
| Vincent et al 2006 [49] | | | | |
| | Canada | Quantitative quasi-experimental design; 975 calls for 38 clients over 6-month period | Positive effect on caregiver burden. Number of home visits by care workers decreased. No improvement in quality of life. | Tele-surveillance |
| Youssef et al 2000 [41] | | | | |
| | United Kingdom | Quantitative study; recorded calls to a control center for 6 months | Total of 542 alarms excluding false alarms. Caregiver solved most problems. GP[e] was called on 38 occasions, ambulance called on 91 occasions, 44 transported to ED, and 29 admitted. | Community alarm |

[a]PERS: personal emergency response system.

[b]RCT: randomized controlled trial.

[c]ED: emergency department.

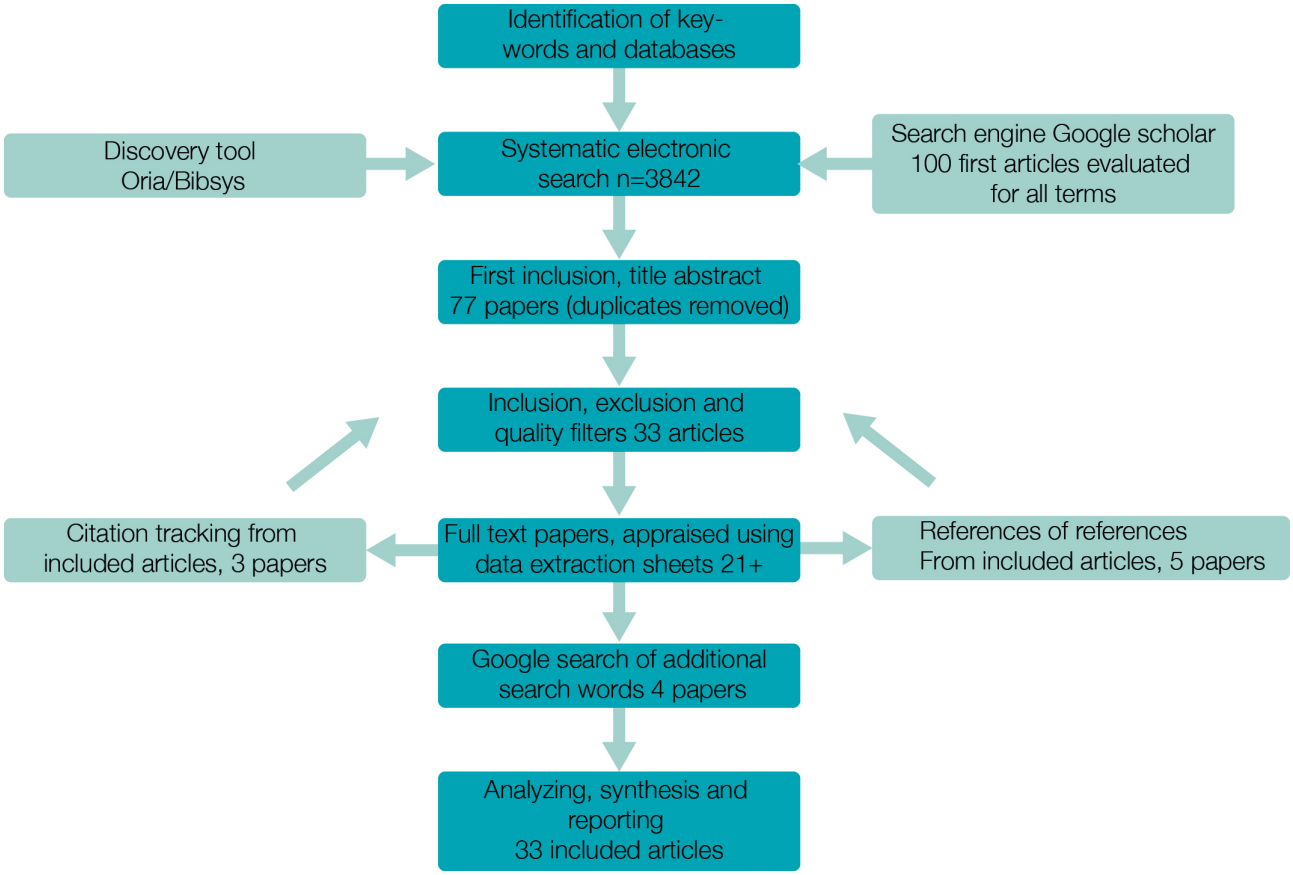
[d]RHQ: reach help quickly.

[e]GP: general practitioner.



XSL•FO
RenderX

**Figure 1.** Overview of the article searching process.




## Results

### Overview

A total of 33 peer-reviewed articles were included, all published between 1987 and 2014 [9,16,22-52]. The articles differ in purpose and study design and represent a wide range of methodology and research traditions. There has been a methodological development in the field. Simplified, descriptive, quantitative evaluation studies of predefined effects dominated the early studies. From the year 2000, we see both qualitative and quantitative studies increasingly displaying a perception of the complexity of the service and society, for example, in-depth views of frail old women's personal experiences with the PERS through the phenomenological studies of Porter et al [27-32].

Table 1 shows that out of 33 studies, 20 (61%) focused directly on the end user, and were related to different aspects of having a PERS. A total of 5 studies (15%) used registered data from the alarm centers regarding use, malfunction, etc. A total of 2 studies (6%) looked mainly at how use of the alarm affects emergency admission, response time, and economy. The remaining studies focused on the service system, the service provider's experience, and the service organization.

As shown in Table 1, different terms are used for the alarm system. The *personal emergency response system (PERS)* is the term most commonly used in articles from the United States, Australia, Germany, the United Kingdom, and Sweden. Other common terms mainly used in the United Kingdom, Scandinavia, and Australia are variations of *personal*, *safety*, *social*, and *community alarms*.

The following section will follow Nicolini's [18] suggestions of "zooming in" on practice, what people do or say, patterns of relationships, and what mechanisms achieve durability in time. This is done by focusing on the included studies' descriptions of the end users' experiences with the PERS, followed by other actors' experiences. Thereafter, I will describe how the included studies describe the interaction between the human actors and the technology, as well as the different actors' wishes for future telecare.

### The End Users' Experiences With the Personal Emergency Response System

Summing up the demographic data from the included studies, the typical PERS end user is an old, fragile woman, living alone, over 80 years of age with physical problems and in need of assistance. The articles state that the end users find the alarm easy to use. Only 2 out of 33 studies (6%) describe demands for training and information and suggest that short learning sessions are preferable to one initial, long session [16,33].

Even though most studies indicate the usefulness of the alarm for fragile elderly people, the alarm does not seem suitable for everybody. Roush and Teasdale [34] found that it is difficult to establish who would utilize a PERS. It is a useful way of getting help faster when the alarm is accepted, understood, and used [35]. A significant proportion of the elderly are less likely to utilize the PERS when in need, especially confused persons [13,36]. Inability to press the button, forgetting to wear it, failing to remember that one is wearing it, or being unable to let helpers inside are reasons given for end users not using the alarm [31,37].



XSL•FO
RenderX

The most stated reason for getting a PERS was the possibility of getting help fast in an emergency [27,30,32,33,35,38-40]. Living in isolation, poor mobility, experiences with hospital stays after a "long lie" following a fall, and concern for personal safety were all catalysts for acquiring the alarm [29,35,37,41].

The articles found that many end users were satisfied with the PERS overall, since it enabled them to summon help if necessary, and that staff were patient despite false or repeated alarms [9,31,33,34,38,42-44]. The studies reported success stories involving activation of the alarm [33]. The ability to get help faster provides a sense of security. Roush et al [40] found that PERS users experienced higher levels of well-being. Although a randomized controlled trial (RCT) assessing the impact of a PERS on anxiety found there was a slightly decreased fear of falling, there was, however, no reduction in anxiety [45]. Another study found that regardless of positive experiences, there was no significant improvement in quality of life [46].

Almost all included studies discussed reasons for activation of the alarm. They demonstrated quite different results regarding the frequency of activation due to emergencies. In one study, only 2.1% of alarm activations were due to emergencies whereas in another, the figure was 67%. Falling was the most common reason for emergency calls, and many users had fallen more than once. Other medical emergencies were also common [33,36,38,43,44,46-51]. End users with a positive experience from a previous emergency were more compliant and satisfied, and men were more likely than women to use the alarm more frequently [33,41,49].

Despite satisfaction with the service, the studies found challenging experiences for the alarm users as well. One study describes fear and insecurity regarding whether the PERS would function when needed, especially at nighttime [42]. The PERS was found to increase ability to live independently [43], and having the alarm was of importance in maintaining end users' lifestyles. Some stated that the PERS helped them to keep their social networks intact [27] and to resume activities they had enjoyed previously [38]. Other studies reported a negative effect on the end user's social life due to uncertainty about the pendant range [37]. To feel safe, the solution was to stay indoors [16,42,52].

Nyman and Victor [41] found that there was not necessarily a correlation between the perception of being satisfied/thinking it is important and actually wearing or using the alarm. They found that the PERS was highly accepted but rarely used. The included studies reported rather different results regarding whether the respondents wore the alarm pendant. End users who considered the PERS important for them wore the pendant significantly more [9,43,47,50]. A total of 2 studies out of 33 (6%) found that about 25% of the respondents never wore the pendant [9,34].

According to the articles, the end users had many reasons for not wearing the pendant: "forgot to put it on," "worry it will get damaged," "do not think they need it at the time," "not satisfied with the PERS," and "uncomfortable to wear" [27,30,45]. Porter [30] found that all women interviewed who wore the pendant did so unwillingly. The PERS made it possible to live alone, but also made life more complicated due to choices as to when to wear and activate the alarm, and fear of triggering it by accident.

It is clear that the alarms are also used for purposes that were not foreseen or intended [33,38,43,45,53,54]. Porter [28] found that some end users would use their PERS if an intruder came, believing that the loud voice when connecting would scare off a burglar. The number of false alarms varied considerably between the studies, and so did what counted as a false alarm.

Several studies found that some users would not activate the PERS even in emergencies [9,30,34,48]. Reasons given were as follows: "wanted to manage on their own," "forgot," "call neighbor," "see if it passes," "don't want to be dragged off to hospital," "afraid to bother," "called 911," "unsure whether serious enough emergency," "don't want strangers in the house," and "unsure of helpers' qualifications" [16,29,32,38,43]. Many respondents never felt in need, and therefore never activated their alarm [47]. There was little focus in the articles on how end users assessed the appearance of the PERS pendant, although there were comments on "stigmatizing" appearance in 5 out of 33 studies (15%) [16,30,38,42,54].

Economic issues were mentioned, mainly in studies from the United States and Canada. Users often paid a fee for having the PERS. Both users and responders raised concerns about the costs and felt it to be too expensive [27,33,49,50]. A total of 2 out of 33 studies (6%) described how some users would like to have the PERS but could not afford it [33,35].

### How Other Actors Experience the Personal Emergency Response System

The included studies focused very little on how relatives experienced the PERS. Some studies mentioned that having the alarm gave families peace of mind and reduced their burden [27,30,33,35,38,40,43]. Studies in which private persons were first responders reported that most were happy to remain so [33,47] with the exception of the study of Sjölinder and Avatare Nöu [52].

The studies described different service organizations, from directly distributing the alarm call to a nominated contact, to larger or smaller private or public response centers. Control center operators' tasks varied according to the service offered. Some response centers were staffed with health care workers [33,47,53,55]. The staff at different call centers had different experiences and attitudes toward the PERS. One large study in Finland found several bottlenecks in the service [55]. Care workers described how the alarm sometimes caused harm and extra work due to accidents, technical failures, and difficulties separating false alarms from emergencies [37]. Experiencing bureaucratic and organizational challenges, they found the alarm to be stressful, costly, and difficult. Others had positive experiences with increased work motivation and better workload planning with reduced visits to end users. The end users gained more privacy and received help only when needed, making night shifts easier [16,56].



XSL•FO
RenderX

### How the Personal Emergency Response System Affects the Interaction Between the Actors

In the included studies, there was little focus on the interactions between the different actors involved in the PERS. Pekkarinen and Melkas [16] found in their study that the holistic situation of end users was not always understood and well-managed by the service providers [16,56]. Some end users reported a less than satisfactory response from the monitoring center and slow response time [16]. The operators were sometimes impolite or the end users' needs were underestimated [49]. Misunderstandings caused by dialects or unclear speech, insufficient follow-up after hospital discharge, and fear of being a burden were problems described [37]. A total of 2 out of 33 studies (6%) found that users expressed fear of causing false alarms by setting the alarm off accidentally, resulting in strangers' voices in their homes [30,31].

Studies described how respondents became motivated to request an alarm in different ways. Health care workers and family were the main source for suggesting a PERS. Respondents became more motivated to use a PERS if health care personnel rather than family suggested this [35,45].

Some articles reported that having the PERS reduced end users' contact with family [43], leaving them feeling lonely, having only the alarm [56]. Both care workers and end users described fears that technology would replace personal service and the support of friends, family, etc. [16,51]. The PERS allowed users to get help when needed, but there was little description of what "when needed" implied. Some studies described the PERS merely as a medical emergency system [30,35,40,43,49,51,54]. Others indicated that PERS is a service that includes guidance in health and medication questions and social calls in addition to being an emergency system [53,54,57].

As reflected in the name, the PERS is a technical device integrated in a service system. Many of the included articles touched on technological problems, even though this is a well-established technology. There is no connection between the technical failure reported and the age of the studies. In addition to limited and confusing alarm range, reported problems included insufficient speaker capability, battery failure, varying needs for button sensitivity, and nonreplaceable parts of the device [16,33,42,47,52,54,55].

### Wishes for the Future

Some studies described wishes for improvement of the PERS. End users wanted longer pendant range, smaller pendants, and for the PERS to be waterproof, personalized, include global positioning system (GPS) and relevant alarms, automatic connection to the nearest health personnel, and automatic dispatching sound when in need [16,42]. They also suggested how service could be improved by responders identifying themselves and speaking slowly and loudly, and that written materials in large print should be provided [33].

## Discussion

### Principal Findings

The key objectives of this integrative review were to explore existing research on the PERS and to seek insight into how actors experience this technology in use in home care services, thus providing a richer and more nuanced view of how actors interact with technologies in caring practices. By following the *theory-method package*, as described by Nicolini [18], the "zooming in" on the practice of the PERS in use as displayed in the Results section will be followed by "zooming out" in this Discussion section, following trails of connections between the PERS in use and other telecare practices. The focus for further discussion is how terms and place matter, and how different actors interact and create changes in roles, use, interactions, and practices. In this way, we can acquire a wider picture of technology in use in caring practices. This provides us with insights of what makes the practice of the PERS so durable over time and contributes to an understanding of what we can draw from this to other caring practices with telecare in use.

### How Words Create Reality

The many different terms used for this technology may reflect different conceptions of the purpose of the technology. The term *personal emergency response system* indicates that the purpose is to respond to an emergency. The term *safety* or *social alarm* indicates that the alarm might include help with social issues and is there for the end user's safety. Technologies are scripted, like a play or a film. This means that the designer and producer have context and users in mind when developing technology [58]. Oudshoorn [20] argues that a large part of the practice involving different kinds of telecare involves filling the gaps between the scripts and the technology practice. This diversity in terms, aim, and purpose seems to result in uncertainty among the end users about what is a legitimate use of the alarm as described in several of the included articles. Thus, this illustrates differences in the ways the script of the technology is presented and lived.

### The Privacy of the Home

Oudshoorn [20] argues that implementing telecare in someone's home creates changes. The home is no longer the same private sphere when connected to health care centers. She describes this as a medicalization of the home. The care personnel only come when the end user for some reason activates the alarm. Even so, the results show how end users' fears of activating the alarm by accident with the subsequent arrival of "strangers" in the house increases their anxiety. Further development of the passive alarm in connection with the third generation of the PERS challenges the definition of a private home even more, turning the home into a place for monitoring health and daily living. Milligan and Wiles [59] describe this as technologies creating "cracks in the door," allowing care personal to monitor and enter the home without physical presence [20,59]. On the other hand, this review shows how health care workers relate that they no longer have to disturb the patient's privacy at night, knowing that he/she will use the alarm if in need. Some end users describe how knowing the connection is there makes them feel safe so that they dare to be more active. Hence, changing



XSL•FO
RenderX

the home from a private place to an arena for telecare in caring practices will change the home in complex and contradictory ways.

### Changing Roles in Caring Practices

Telecare promises an opportunity to get help when needed in one's home. Previous studies showed little reference to the shift this brings about in the redistribution of responsibility among the actors involved, with the delegation of major responsibility to care workers and the end user [21,24]. Use of telecare redefines the patient role from a passive recipient to actively participating in health care and safety monitoring, and demands that patients become competent users of the technology [20]. This demand for active participation largely delegates the responsibility for their own safety to end users. The results of this review found that the PERS, even though it is considered easy to use, is not for everybody. The PERS gives the end user an active role in their care. He/she has to press the pendant to reach help. The results describe how end users do not always remember or manage to activate the alarm. The PERS is therefore not suited for end users who are unable to activate the alarm when in need, for instance, mentally confused people.

This responsibility is further extended in devices developed for the third generation of the PERS, where patients are expected to some extent to manage and monitor their own health. This is a double-edged sword: patients on the one hand gain knowledge and ownership of their own health, but often face demanding requirements to master the technology within the strict frames decided by the system and the technology. This requires that the end user develop skills in using the technology in possibly stressful situations since the caregiver is not present in the home.

These changes in roles, responsibilities, and work were largely disregarded in the studies included in this review. The findings paid little attention to the changes in the role of the care personnel. However, this review reveals a diversity and complexity in health care workers' experiences with the PERS, varying from finding that the technology enables them to give the users more privacy and freedom, to complicating the organization of the work and causing stress and failure in the caring practice.

### Resistance and Nonuse of Telecare

Promoters of technology innovations tend to describe resistance and nonuse of technology due to lack of technological skills and access among older people as a generational issue [20,60]. Akrich and Latour found that instead of complaining about the technology, actors tend to adjust their practices or resist using the technology [20,58,61].

Despite the main finding of the studies included in this review of users showing great satisfaction with the PERS, the results described how many users acquire the alarm, but hardly ever wear or activate the alarm pendant. As previously described, it seems clear that some nonuse described in the results was related to lack of ability to utilize the alarm, but that is not the whole story. Those not using the alarm often had relevant reasons for doing so. In addition to the previously mentioned reasons, many end users described how they did not feel they needed the PERS, found other solutions for being safe, did not want to bother or be bothered, or found the PERS stigmatizing. It seems that nonuse is more complex than the users' lack of skills and access, though that is also important. This review shows that resistance and nonuse are due to factors such as the change in caring practices and the way users experience the technology as changing their lives and homes. The results also indicate how end users experience challenges related to use of the PERS, including technological failure, fear of the alarm not working, and limited alarm pendant range.

Pols [15] affirmed that different user groups of telecare tend to be defined by similarities within the group, but she stresses that there are huge differences and heterogeneities within different groups. The studies included in this review described the end users of the alarm as fragile, high-dependence older people, often with an extensive medical history and often living alone; this indicates a need for the PERS. None of the studies discussed diversity within this group of end users, although some described how respondents talk about resistance to being considered "one of those." It seems that having the PERS defines the end users as part of the group of frail, old, dependent people, and this causes resistance among some.

### How Telecare Creates New Interactions and Practices

Telecare implies a different kind of care with complex interactions between multiple actors and a wide variety of technology and changed roles, thus redefining how actors live, work, and even identify their lives [20]. Results show that having the PERS affected the users in different ways socially. Some became more active because they felt safe having the alarm; others, however, felt restricted by the pendant range and therefore stayed inside their homes. Some actors feared that the alarm might replace human contact.

This review reveals a huge diversity in the experiences and roles of the actors involved, presenting a variety of experiences, both negative and positive, from end users, care personnel, and other actors. Therefore, we cannot really talk about one type of practice related to the PERS, but rather a variety of practices as a result of the interactions between the technology and the actors involved, and how the service is organized and carried out.

The history of integrating telecare in community care shows that the technology in use tends to work in unforeseen and different ways than intended [20]. The results of this review show how the end users activated the alarm for a number of reasons, and found new and other functions for the alarm than the one intended (eg, older women planning to use the PERS to scare away unwanted intruders). Users also had wishes for future functions that would increase the value of having the PERS, for example, increased pendant range, integrated GPS, and smaller pendants, to mention a few.

### What the Technology Does—and What We Think it Does

Pols [15] found in her study of telecare in the Netherlands that use of telecare did not solve existing practice problems but was instrumental in creating new practices with different challenges and problems, and thereby changed the actors' lives.



XSL•FO
RenderX

The PERS is introduced as technology that makes users safe in their homes and enables them to reach help when needed. This review shows how end users in the included studies expressed satisfaction with the technology, and experienced well-being and a sense of security. However, having the PERS did not reduce anxiety or improve their quality of life. Some of the studies included in this review found that having the PERS did reduce hospital days and medical complications due to long lies after falls. Moreover, the material described success stories involving activating the alarm. To some extent, the alarm thereby fulfilled its promise of increasing safety at home. However, what the results show is that the picture is much more complicated than the PERS simply being an easy fix for anxiety and risk experienced by frail older people living alone.

## Conclusions

This review reveals how rather simple and well-established telecare technologies such as the PERS are actually complex, integrated caring practices that interact with the different actors involved and create changes in daily living.

The PERS has proven to be durable over time, while many telecare technologies tend never to leave the pilot stage. The reasons for this are complicated, but the results describe some contributing factors. Many users find the PERS to be easy to use and it makes it possible for the end users to live independently by providing help and safety when needed, giving the end users an active role in the caring practice. The PERS in many ways delivers its promises of safety and independent living.

While the Results section describes how the PERS contributes to safety and independence and discusses what the PERS means for the actors involved, it also reveals unforeseen consequences of the alarm and possible improvements in both the device and the service. This review provides us with an understanding of the complexity of practice by showing how even rather simple technology interacts with actors and redefines how they live and work, and even how the technology affects their identities. The Discussion section problematizes this by "zooming out" and argues for an approach to telecare in which the complexity of practice is accounted for, where actors' resources, attitudes, and abilities are considered when choosing technology.

This paper shows how technology, involved actors, network, and context must be thought of together as part of practice. This calls for a sensitivity to what it means for involved actors when we redistribute responsibility to the end users, and change the roles and work practices of the caring personnel. Another key factor is taking into account how implementing telecare changes the idea of home and all it represents for the actors.

There is a need to be sensitive to diversity in apparently homogenous groups when adopting new telecare technologies in home care practices, and to acknowledge that technology is never neutral. This review shows how understanding end users' experiences is an important resource for understanding how technology innovations in caring practices are actants in creating new caring practices, thus acknowledging that there are many reasons for resisting and failing to use the technology.

It is time to move away from thinking of telecare technologies as black boxes that can be implemented without changing the caring practice. This review shows how the plug-and-play expectations producers tend to generate is a simplification of the reality. It seems clear that "one size doesn't fit all." This calls for a recognition that place and actors matter, and a sensitivity for the practices in which the technology is adopted is necessary.

## Acknowledgments

Sincere thanks to supervisors Professor Lars Fuglsang, Professor Ragnhild Hellesø, Professor Rolf Rønning, Maren Sogstad, PhD, and staff at Norwegian University of Science and Technology (NTNU) in Gjøvik Library.

## Conflicts of Interest

None declared.

## Multimedia Appendix 1

Thorough description of the research strategy process.

[PDF File (Adobe PDF File), 17KB - jmir_v18i7e187_app1.pdf ]

## Multimedia Appendix 2

Methodological considerations.

[PDF File (Adobe PDF File), 18KB - jmir_v18i7e187_app2.pdf ]

## References

1. Melander-Wikman A, Fältholm Y, Gard G. Safety vs. privacy: Elderly persons' experiences of a mobile safety alarm. Health Soc Care Community 2008 Jul;16(4):337-346. [doi: 10.1111/j.1365-2524.2007.00743.x] [Medline: 18613909]
2. Tarricone R, Tsouros AD, editors. Home Care in Europe: The Solid Facts. Geneva, Switzerland: World Health Organization; 2008.



XSL•FO
RenderX

3. Gasior K, Huber M, Lamura G, Lelkes O, Marin B, Rodrigues R, et al. In: Rodrigues R, Huber M, Lamura G, editors. Facts and Figures on Healthy Ageing and Long-Term Care: Europe and North America. Vienna, Austria: European Centre for Social Welfare Policy and Research; 2009.
4. Boudiny K. 'Active ageing': From empty rhetoric to effective policy tool. Ageing Soc 2013 Aug;33(6):1077-1098 [FREE Full text] [doi: 10.1017/S0144686X1200030X] [Medline: 23913994]
5. Foster L, Walker A. Gender and active ageing in Europe. Eur J Ageing 2013;10:3-10. [doi: 10.1007/s10433-013-0261-0]
6. Halford S, Lotherington AT, Obstfelder A, Dyb K. Getting the whole picture? Inf Commun Soc 2010;13(3):442-465. [doi: 10.1080/13691180903095856]
7. Andreassen HK, Kjekshus LE, Tjora A. Survival of the project: A case study of ICT innovation in health care. Soc Sci Med 2015 May;132:62-69. [doi: 10.1016/j.socscimed.2015.03.016] [Medline: 25795426]
8. Sundbo J, Toivonen M, editors. User-Based Innovation in Services. Cheltenham, UK: Edward Elgar Publishing; 2011.
9. Heinbüchner B, Hautzinger M, Becker C, Pfeiffer K. Satisfaction and use of personal emergency response systems. Z Gerontol Geriatr 2010 Aug;43(4):219-223. [doi: 10.1007/s00391-010-0127-4] [Medline: 20814795]
10. Fisk MJ. Social Alarms to Telecare: Older People's Services in Transition. Bristol, UK: Policy Press; 2003.
11. Mørk E. Kommunale helse- og omsorgstjenester 2014 tatistikk om tjenester og tjenestemottakere. In: Oslotatistisk Sentralbyrå. Oslo, Norway: Statistisk Sentralbyrå; 2015.
12. Mayhew-Rankcom VJ, Lindeman MA, Hill KD, Smith RA. Who should get personal alarms? The development of instruments for the assessment of need. Aust J Prim Health 2001;7(3):54-62. [doi: 10.1071/PY01047]
13. Sugarhood P, Wherton J, Procter R, Hinder S, Greenhalgh T. Technology as system innovation: A key informant interview study of the application of the diffusion of innovation model to telecare. Disabil Rehabil Assist Technol 2014 Jan;9(1):79-87. [doi: 10.3109/17483107.2013.823573] [Medline: 23919407]
14. Pritchard GW, Brittain K. Alarm pendants and the technological shaping of older people's care: Between (intentional) help and (irrational) nuisance. Technol Forecast Soc Change 2015 Apr;93:124-132.
15. Pols J. Telecare: What patients care about. In: Mol A, Moser I, Pols J, editors. Care in Practice: On Tinkering in Clinics, Homes and Farms. Bielefeld, Germany: Transcript Verlag; 2010:171-194.
16. Pekkarinen S, Melkas H. Safety alarm systems and related services: From potholes to innovation opportunities. Int J Serv Sci Manag Eng Technol 2010;1(4):53-70. [doi: 10.4018/jssmet.2010100105]
17. Sanchez-Criado T, Lopez D, Roberts C, Domenech M. Installing telecare, installing users: Felicity conditions for the instauration of usership. Sci Technol Human Values 2014 Jan 09;39(5):694-719. [doi: 10.1177/0162243913517011]
18. Nicolini D. Practice Theory, Work, and Organization: An Introduction. Oxford, UK: Oxford University Press; 2012.
19. Oudhoorn N, Pinch T. User-technology relationships: Some recent developments. In: Hackett EJ, Amsterdamska O, Lynch M, Wajcman J, editors. The Handbook of Science and Technology Studies. 3rd edition. Cambridge, MA: The MIT Press; 2008:541-566.
20. Oudshoorn N. Telecare Technologies and the Transformation of Healthcare. Hampshire, UK: Palgrave MacMillan; 2011.
21. Oudshoorn N. How places matter: Telecare technologies and the changing spatial dimensions of healthcare. Soc Stud Sci 2012 Jan 16;42(1):121-142. [doi: 10.1177/0306312711431817]
22. Pols J. Care at a Distance: On the Closeness of Technology. Amsterdam, the Netherlands: Amsterdam University Press; 2012.
23. Pols J, Willems D. Innovation and evaluation: Taming and unleashing telecare technology. Sociol Health Illn 2011 Mar;33(3):484-498 [FREE Full text] [doi: 10.1111/j.1467-9566.2010.01293.x] [Medline: 21241338]
24. Nicolini D. The work to make telemedicine work: A social and articulative view. Soc Sci Med 2006 Jun;62(11):2754-2767. [doi: 10.1016/j.socscimed.2005.11.001] [Medline: 16343724]
25. Nicolini D, Gherardi S, Yanow D, editors. Knowing in Organizations: A Practice-Based Approach. Armonk, NY: ME Sharpe; 2003.
26. Whittemore R, Knafl K. The integrative review: Updated methodology. J Adv Nurs 2005 Dec;52(5):546-553. [doi: 10.1111/j.1365-2648.2005.03621.x] [Medline: 16268861]
27. Porter EJ. Moments of apprehension in the midst of a certainty: Some frail older widows' lives with a personal emergency response system. Qual Health Res 2003 Nov;13(9):1311-1323. [Medline: 14606415]
28. Porter EJ. Contemplating what I would do if someone got in my house: Intentions of older homebound women living alone. ANS Adv Nurs Sci 2008;31(2):106-118. [doi: 10.1097/01.ANS.0000319561.84007.f9] [Medline: 18497587]
29. Porter EJ, Lasiter S. Older homebound women's perceived risk of being unable to reach help quickly: Influence of situations involving age-peers. Res Nurs Health 2012 Dec;35(6):624-633 [FREE Full text] [doi: 10.1002/nur.21507] [Medline: 23070958]
30. Porter EJ. Wearing and using personal emergency respone system buttons. J Gerontol Nurs 2005 Oct;31(10):26-33. [Medline: 16262088]
31. Porter EJ, Ganong LH. Considering the use of a personal emergency response system: An experience of frail, older women. Care Manag J 2002;3(4):192-198. [Medline: 12847936]
32. Porter EJ, Ganong LH, Matsuda S. Intentions of older homebound women with regard to reaching help quickly. West J Nurs Res 2013 Jan;35(1):6-23. [doi: 10.1177/0193945911428482] [Medline: 22146885]




XSL•FO
RenderX

33. Fallis WM, Silverthorne D, Franklin J, McClement S. Client and responder perceptions of a personal emergency response system: Lifeline. Home Health Care Serv Q 2007;26(3):1-21. [doi: 10.1300/J027v26n03_01] [Medline: 17804350]
34. Roush RE, Teasdale TA. Personal emergency response services: Do the benefits justify the cost in seniors housing and care properties? Seniors Hous Care J 2011;19(1):73-82.
35. Johnston K, Grimmer-Somers K, Sutherland M. Perspectives on use of personal alarms by older fallers. Int J Gen Med 2010;3:231-237 [FREE Full text] [Medline: 20830199]
36. Tinker A. Alarms and telephones in emergency response--Research from the United Kingdom. Home Health Care Serv Q 1992;13(3-4):177-189. [Medline: 10126440]
37. Johnston K, Worley A, Grimmer-Somers K, Sutherland M, Amos L. Personal alarm use to call the ambulance after a fall in older people: Characteristics of clients and falls. Australas J Paramed 2010;8(4):1-9.
38. Fisk MJ. A comparison of personal response services in Canada and the UK. J Telemed Telecare 1995;1(3):145-156. [Medline: 9375135]
39. Olsson A, Engström M, Skovdahl K, Lampic C. My, your and our needs for safety and security: Relatives' reflections on using information and communication technology in dementia care. Scand J Caring Sci 2012 Mar;26(1):104-112. [doi: 10.1111/j.1471-6712.2011.00916.x] [Medline: 21843198]
40. Roush RE, Teasdale TA, Murphy JN, Kirk MS. Impact of a personal emergency response system on hospital utilization by community-residing elders. South Med J 1995 Sep;88(9):917-922. [Medline: 7660208]
41. Nyman S, Victor C. Use of personal call alarms among community-dwelling older people. Ageing Soc 2014;34(1):67-89.
42. Boström M, Kjellström S, Malmberg B, Björklund A. Personal emergency response system (PERS) alarms may induce insecurity feelings. Gerontechnology 2011;10(3):140-145. [doi: 10.4017/gt.2011.10.3.001.00]
43. De San Miguel K, Lewin G. Personal emergency alarms: What impact do they have on older people's lives? Australas J Ageing 2008 Jun;27(2):103-105. [doi: 10.1111/j.1741-6612.2008.00286.x] [Medline: 18713202]
44. Hyer K, Rudick L. The effectiveness of personal emergency response systems in meeting the safety monitoring needs of home care clients. J Nurs Adm 1994 Jun;24(6):39-44. [Medline: 8006702]
45. Lee JS, Hurley MJ, Carew D, Fisher R, Kiss A, Drummond N. A randomized clinical trial to assess the impact on an emergency response system on anxiety and health care use among older emergency patients after a fall. Acad Emerg Med 2007 Apr;14(4):301-308 [FREE Full text] [doi: 10.1197/j.aem.2006.11.017] [Medline: 17331915]
46. Youssef GG, Underhill TJ, Tovey C. Referral to the accident and emergency department following the use of community alarms. J Accid Emerg Med 2000 Sep;17(5):348-350 [FREE Full text] [Medline: 11005406]
47. Farquhar DL, Fonda D, Danek CL, Ryan AM. A survey of personal emergency alarms for the frail elderly. Australas J Ageing 1992;11(1):3-7. [doi: 10.1111/j.1741-6612.1992.tb00529.x] [Medline: 17416612]
48. Fleming J, Brayne C, Cambridge City over-75s Cohort (CC75C) study collaboration. Inability to get up after falling, subsequent time on floor, and summoning help: prospective cohort study in people over 90. BMJ 2008;337:a2227 [FREE Full text] [Medline: 19015185]
49. Levine DA, Tideiksaar R. Personal emergency response systems: Factors associated with use among older persons. Mt Sinai J Med 1995 Sep;62(4):293-297. [Medline: 7565852]
50. Mann W, Belchior P, Tomita M, Kemp B. Use of personal emergency response systems by older individuals with disabilities. Assist Technol Spr 2005;17(1):82-88. [doi: 10.1080/10400435.2005.10132098] [Medline: 10400435]
51. McWhirter M. A dispersed alarm system for the elderly and its relevance to local general practitioners. J R Coll Gen Pract 1987 Jun;37(299):244-247 [FREE Full text] [Medline: 3448211]
52. Sjölinder M, Avatare NA. Indoor and outdoor social alarms: Understanding users' perspectives. JMIR Mhealth Uhealth 2014;2(1):e9 [FREE Full text] [doi: 10.2196/mhealth.2730] [Medline: 25099060]
53. Premik M, Rudel D, Licer N. Could the community social alarm system serve also as a primary health care service? Med Arh 1998;52(3):153-155. [Medline: 9863321]
54. Vincent C, Reinharz D, Deaudelin I, Garceau M, Talbot LR. Public telesurveillance service for frail elderly living at home, outcomes and cost evolution: A quasi experimental design with two follow-ups. Health Qual Life Outcomes 2006;4:41 [FREE Full text] [doi: 10.1186/1477-7525-4-41] [Medline: 16827929]
55. Melkas H. Network collaboration within safety telephone services for ageing people in Finland. Gerontechnology 2003;2(4):306-323. [doi: 10.4017/gt.2003.02.04.003.00]
56. Melkas H. Informational ecology and care workers: Safety alarm systems in Finnish elderly-care organizations. Work 2010;37(1):87-97. [doi: 10.3233/WOR-2010-1060] [Medline: 20858991]
57. Raappana A, Rauma M, Melkas H. Impact of safety alarm systems on care personnel. Gerontechnology 2007;6(2):112-117.
58. Akrich M. The de-scription of technical objects. In: Bijker WE, Law J, editors. Shaping Technology/Building Society: Studies in Sociotechnical Change. Cambridge, MA: The MIT Press; 1992:205-224.
59. Milligan C, Wiles J. Landscapes of care. Prog Hum Geogr 2010 Apr 23;34(6):736-754. [doi: 10.1177/0309132510364556]
60. Wyatt S. Non-users also matter: The construction of users and non-users of the Internet. In: Oudshoorn N, Pinch T, editors. How Users Matter: The Co-Construction of Users and Technology. Cambridge, MA: The MIT Press; 2003:67-80.




XSL•FO
RenderX

61. Akrich M, Latour B. A summary of a convenient vocabulary for the semiotics of human and nonhuman assemblies. In: Bijker WE, Law J, editors. Shaping Technology/Building Society: Studies in Sociotechnical Change. Cambridge, MA: The MIT Press; 1994:259-264.

## Abbreviations

**ED:** emergency department
**GP:** general practitioner
**GPS:** global positioning system
**NTNU:** Norwegian University of Science and Technology
**PERS:** personal emergency response system
**RCT:** randomized controlled trial
**RHQ:** reach help quickly
**STS:** science and technology studies

*Edited by G Eysenbach; submitted 08.03.16; peer-reviewed by M Sjölinder, H Andreassen; comments to author 31.03.16; revised version received 15.05.16; accepted 11.06.16; published 14.07.16.*

*Please cite as:*
*Stokke R*
*The Personal Emergency Response System as a Technology Innovation in Primary Health Care Services: An Integrative Review*
*J Med Internet Res 2016;18(7):e187*
*URL: http://www.jmir.org/2016/7/e187/*
*doi:10.2196/jmir.5727*
*PMID:27417422*

©Randi Stokke. Originally published in the Journal of Medical Internet Research (http://www.jmir.org), 14.07.2016. This is an open-access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/2.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on http://www.jmir.org/, as well as this copyright and license information must be included.



XSL•FO
RenderX

Original Paper

# Behavioral Analysis of Visitors to a Medical Institution's Website Using Markov Chain Monte Carlo Methods

Teppei Suzuki[1], MHS, MBA; Yuji Tani[2], PhD, MBA; Katsuhiko Ogasawara[1], PhD, MBA

[1]Faculty of Health Sciences, Graduate School of Health sciences, Hokkaido University, Sapporo, Japan

[2]Department of Medical Informatics and Hospital Management, Asahikawa Medical University, Asahikawa, Hokkaido, Japan

**Corresponding Author:**
Katsuhiko Ogasawara, PhD, MBA
Faculty of Health sciences
Graduate School of Health Sciences
Hokkaido University
Kita12, Nishi 5, Kita-ku
Sapporo,
Japan
Phone: 81 11 706 3409
Fax: 81 11 706 2851
Email: oga@hs.hokudai.ac.jp

## Abstract

**Background:** Consistent with the "attention, interest, desire, memory, action" (AIDMA) model of consumer behavior, patients collect information about available medical institutions using the Internet to select information for their particular needs. Studies of consumer behavior may be found in areas other than medical institution websites. Such research uses Web access logs for visitor search behavior. At this time, research applying the patient searching behavior model to medical institution website visitors is lacking.

**Objective:** We have developed a hospital website search behavior model using a Bayesian approach to clarify the behavior of medical institution website visitors and determine the probability of their visits, classified by search keyword.

**Methods:** We used the website data access log of a clinic of internal medicine and gastroenterology in the Sapporo suburbs, collecting data from January 1 through June 31, 2011. The contents of the 6 website pages included the following: home, news, content introduction for medical examinations, mammography screening, holiday person-on-duty information, and other. The search keywords we identified as best expressing website visitor needs were listed as the top 4 headings from the access log: clinic name, clinic name + regional name, clinic name + medical examination, and mammography screening. Using the search keywords as the explaining variable, we built a binomial probit model that allows inspection of the contents of each purpose variable. Using this model, we determined a beta value and generated a posterior distribution. We performed the simulation using Markov Chain Monte Carlo methods with a noninformation prior distribution for this model and determined the visit probability classified by keyword for each category.

**Results:** In the case of the keyword "clinic name," the visit probability to the website, repeated visit to the website, and contents page for medical examination was positive. In the case of the keyword "clinic name and regional name," the probability for a repeated visit to the website and the mammography screening page was negative. In the case of the keyword "clinic name + medical examination," the visit probability to the website was positive, and the visit probability to the information page was negative. When visitors referred to the keywords "mammography screening," the visit probability to the mammography screening page was positive (95% highest posterior density interval = 3.38-26.66).

**Conclusions:** Further analysis for not only the clinic website but also various other medical institution websites is necessary to build a general inspection model for medical institution websites; we want to consider this in future research. Additionally, we hope to use the results obtained in this study as a prior distribution for future work to conduct higher-precision analysis.

***(J Med Internet Res 2016;18(7):e199)*** doi:10.2196/jmir.5139

**KEYWORDS**

information-seeking behavior; Internet; media, social; Bayesian analysis; Web marketing



XSL•FO
RenderX

## Introduction

To reduce the existing "asymmetry of information" between a patient and physician, patients routinely access Web-based information about their health problems and medical treatment options. Internet-based medical resources are constantly being developed and expanded on such an environment [1]. Consistent with the "attention, interest, desire, memory, action (AIDMA)" model of consumer behavior, patients collect information about the medical institutions available using the Internet to select information for their particular needs [2].

Because of this situation, recently, many medical institutions are intent on improving their websites. With the development in Internet environment and devices, we are now able to obtain information on many medical institutions in diverse ways. Many companies achieve greater advertising effects by active release of information on the Internet. Therefore, medical institutions actively using social media are also increasing. However, such actions and study are insufficiently advanced in Japan.

Market researchers and social psychologists routinely conduct various consumer behavior analyses based on the AIDMA model to predict factors influencing consumer action and purchase decisions and clarify consumer psychology and internal states [3].

Studies of consumer behavior may be found in areas other than medical institution websites. After the expansion of Web advertisements and publicity, Internet sales greatly increased the rate of Internet usage. Research can be conducted using the Web access logs on visitor search behavior. To investigate product sales over the Internet, searching behavior models such as the "search keyword" and "page view" assume that the searches are an expression of consumer needs. In this study, we determined the probability of visits to a certain webpage by the search keyword using the Markov Chain Monte Carlo (MCMC) methods [4,5]. Recently, marketing research has yielded positive results applying Bayesian statistics with improvements in computer count ability and expects to apply them to a greater degree in the future [6-8]. At this time, research applying the patient searching behavior model to medical institution website visitors is lacking.

In this study, we have developed a hospital website search behavior model using a Bayesian approach to clarify the behavior of medical institution website visitors and determine the probability of their visits classified by the search keyword.

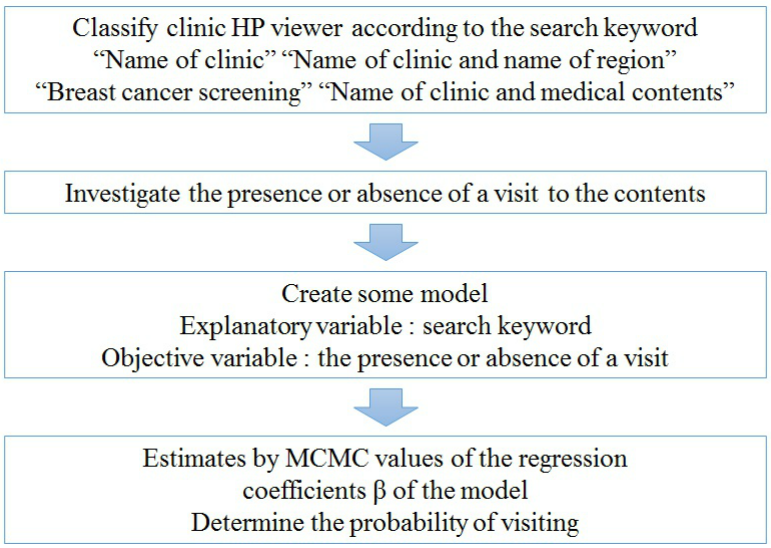
## Methods

### Subject

The flowchart we propose for our research is shown in Figure 1.

We used the website data access log of a clinic of internal medicine and gastroenterology in the Sapporo suburbs for our research, collecting data (336 cases) from January 1 through June 31, 2011. We used Google Analytics to analyze the data access log [9]. The contents of the 6 website pages included the following: home, news, content introduction for medical examinations, mammography screening, holiday person-on-duty information, and other. We used all pages in the clinic for this study. The other page introduces the communication space attached to a hospital. A second visit to the website during the same visit session, distinguished from the first visit and to be counted more correctly as an index page, we classified as "the website (again)." The search keywords we identified as best expressing website visitor needs were listed as the top 4 headings from the access log: clinic name, clinic name + regional name, clinic name + medical examination, and mammography screening.

**Figure 1.** Flowchart.




### *Methods of Analysis*

In this study, we applied Bayes' theorem as the analysis method. The obtained data were y, and the parameter was defined as θ. Both were random variables and are expressed using Bayes' theorem as shown in Figure 2.

The left-hand side was called the posterior distribution. This represented the distribution of θ when data y were obtained.



XSL•FO
RenderX

The right-hand side of $f(y|\theta)$ was the likelihood, and $f(\theta)$ was the distribution of θ. This distribution was called the prior distribution. The distribution of the data expressed by the following equation was represented by $f(y)$. We analyzed using this method as shown in Figure 3.

Using the search keyword as the explaining variable, we built a binomial probit model allowing the inspection of the contents of each purpose variable [10]. The binomial probit model is a discrete selection model used in marketing science. In our study, this model used the formulas as described in Figure 4.

The discrete selection model was formulated to address the behaviors of individuals choosing alternatives from their selection sets. In marketing science, this concept is applied to verify consumer selection behavior [11].

Using this model, we determined the beta value and generated a posterior distribution, showing the visit probability to each category classified by the search keywords.

We performed the simulation using MCMC with a noninformation prior distribution for this model and determined the visit probability classified by keyword for each category. We used the Gibbs sampling method, sampling 50,000 times. We also canceled the first 5000 samples, as an initial dependence period (burn-in) [12].

The joint distribution is expressed as shown in Figure 5.

Generally, to check the convergence of the sampling, autocorrelation function (ACF) is used. Thus, in this study, we used ACF to check sample convergence. With the vertical axis as the autocorrelation coefficient, when autocorrelation is high, the accuracy of the Markov chain is low.

Although the form of the ACF in the determined posterior distribution and convergence was observed, there was a problem in reproducibility. In this research, the log judged precedence research to reference completed by 30 or more and the auto correlation coefficient or less by 0.1. In this research, we used statistical software R (version 2.13.0) for the simulation analysis [12,13].

**Figure 2.** Bayes' theorem.

$$f(\theta|y) = \frac{f(y|\theta)f(\theta)}{f(y)}$$

**Figure 3.** Distribution of the data -f(y).

$$f(y) = \int f(y|\theta)f(\theta)d\theta$$

**Figure 4.** Binomial probit model.

$$y_i = x_i\beta + \varepsilon_i$$

$y_i$ : Existence of a search keyword (yes : 1, no : 0)

$x_i$ : Existence of each page visit (yes : 1, no : 0)

β : Parameter (noninformation prior distribution $\theta - N(0,1 \times 10^3)$)

$\varepsilon_i$ : Error term ($\varepsilon_i - i.i.d.N(0,1)$)

**Figure 5.** Joint distribution.

$$p(y, \beta|X) = p(y|\beta, X)p(\beta)$$

### *Definition of Visit Probability*

To evaluate the posterior probability density function presumed by MCMC, we used the highest posterior density (HPD) interval. As the value was computed using the Bayesian approach and one of the point estimates, the value alone was not sufficient for evaluation purposes. Therefore, for the interval estimate, because all HPDs of the obtained frequency function were either positive or negative, we assumed this was significant and defined the median of the HPD as the visit probability to each category [14]. HPD is not the same as the probability; it may become larger than 1 or less than −1 in value.

## Results

The statistical results for each keyword are shown in Tables 1-4 and Figures 6-9.

When a visitor referred to the keyword “clinic name,” the HPDs to the main page, the website (again), and the contents page were positive. When a visitor referred to the keyword “clinic name + regional name,” the HPDs to the website (again) and the mammography screening page were negative. When a visitor referred to the keyword “clinic name + medical examination,” the HPD to the main page was positive and that to the information page was negative. When a visitor referred to the keyword “mammography screening,” the HPD to the mammography screening was positive.



XSL•FO
RenderX

**Table 1.** Posterior distribution presumption result by keyword “clinic name”.

| | | | 95% HPD[a] interval | | | |
|---|---|---|---|---|---|---|
| Contents | Posterior mean | SD[b] | 2.50% | Median | 97.50% | Convergence |
| Top page | 0.85 | 0.29 | 0.3 | 0.85 | 1.43 | ○ |
| Top page (again) | 0.54 | 0.17 | 0.2 | 0.54 | 0.87 | ○ |
| News | 0.12 | 0.41 | −0.68 | 0.11 | 0.93 | ○ |
| Contents | 0.48 | 0.17 | 0.14 | 0.47 | 0.81 | ○ |
| Mammography screening | 0.03 | 0.15 | −0.26 | 0.04 | 0.33 | ○ |
| Information | 0.32 | 0.17 | −0.01 | 0.32 | 0.65 | ○ |
| Holiday duty hospital | −0.27 | 0.19 | −0.65 | −0.27 | 0.11 | ○ |
| Others | 0.27 | 0.2 | −0.12 | 0.28 | 0.68 | ○ |

[a]HPD: highest posterior density.

[b]SD: standard deviation.

**Table 2.** Posterior distribution presumption result by keyword “clinic name + regional name”.

| | | | 95% HPD[a] interval | | | |
|---|---|---|---|---|---|---|
| Contents | Posterior mean | SD[b] | 2.50% | Median | 97.50% | Convergence |
| Top page | 0.31 | 0.28 | −0.23 | 0.3 | 0.85 | ○ |
| Top page (again) | −0.48 | 0.17 | −0.81 | −0.48 | −0.15 | ○ |
| News | 0.33 | 0.37 | −0.38 | 0.33 | 1.05 | ○ |
| Contents | −0.23 | 0.16 | −0.55 | −0.23 | 0.1 | ○ |
| Mammography screening | −0.5 | 0.15 | −0.79 | −0.5 | −0.22 | ○ |
| Information | 0.01 | 0.16 | −0.31 | 0.01 | 0.32 | ○ |
| Holiday duty hospital | 0.09 | 0.19 | −0.27 | 0.09 | 0.46 | ○ |
| Others | −0.06 | 0.2 | −0.46 | −0.06 | 0.34 | ○ |

[a]HPD: highest posterior density.

[b]SD: standard deviation.

**Table 3.** Posterior distribution presumption result by keyword “clinic name + medical examination”.

| | | | 95% HPD[a] interval | | | |
|---|---|---|---|---|---|---|
| Contents | Posterior mean | SD[b] | 2.50% | Median | 97.50% | Convergence |
| Top page | 17.01 | 9.61 | 1.26 | 15.96 | 37.08 | ○ |
| Top page (again) | −0.02 | 0.45 | −0.92 | −0.01 | 0.86 | ○ |
| News | 13.94 | 12.05 | −0.59 | 10.49 | 40.41 | ○ |
| Contents | 0.01 | 0.41 | −0.78 | 0 | 0.85 | ○ |
| Mammography screening | −0.41 | 0.4 | −1.24 | −0.39 | 0.34 | ○ |
| Information | −0.88 | 0.43 | −1.80 | −0.86 | −0.09 | × |
| Holiday duty hospital | 0.25 | 0.48 | −0.71 | 0.25 | 1.17 | ○ |
| Others | 0.03 | 0.56 | −1.17 | 0.06 | 1.06 | ○ |

[a]HPD: highest posterior density.

[b]SD: standard deviation.



XSL•FO
RenderX

**Table 4.** Posterior distribution presumption result by keyword “mammography screening”.

| | | | 95% HPD[a] interval | | | |
|---|---|---|---|---|---|---|
| Contents | Posterior mean | SD[b] | 2.50% | Median | 97.50% | Convergence |
| Top page | −2.34 | 0.51 | −3.35 | −2.33 | −1.37 | ○ |
| Top page (again) | −0.27 | 0.4 | −1.07 | −0.27 | 0.51 | ○ |
| News | −0.29 | 0.75 | −1.79 | −0.29 | 1.17 | × |
| Contents | −0.5 | 0.33 | −1.16 | −0.5 | 0.14 | ○ |
| Mammography screening | 14.84 | 6.59 | 3.38 | 14.58 | 26.66 | × |
| Information | −0.71 | 0.35 | −1.41 | −0.71 | −0.05 | × |
| Holiday duty hospital | 0.41 | 0.4 | −0.37 | 0.41 | 1.22 | ○ |
| Others | −0.43 | 0.46 | −1.38 | −0.42 | 0.42 | ○ |

[a]HPD: highest posterior density.

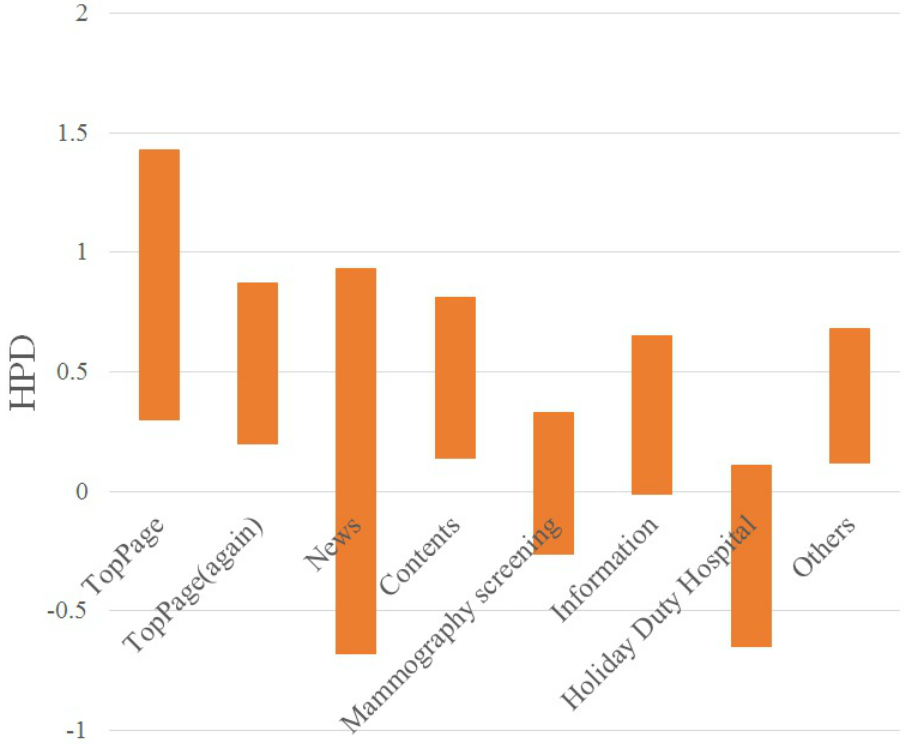
[b]SD: standard deviation.

Next, we showed the results of the simulations regarding the time it took for a visitor to refer to the keyword “clinic name.” Figure 10 shows the presumed posterior distribution, and the horizontal axis is the value of parameter beta. The vertical axis is probability density. Posterior distribution obtained from this simulation was a unimodal distribution. The posterior distribution obtained by the vertical axis expressing probability density was a unimodal distribution.

Figure 11 shows the sampling convergence by MCMC, with the vertical axis as the beta value and the horizontal axis as the sampling number.

Figure 12 shows the ACF obtained by the simulation. For the keyword “clinic name,” the autocorrelation was small, and it was fully completed [15].

**Figure 6.** HPD of posterior distribution by keyword "clinic name".






XSL•FO
RenderX

**Figure 7.** HPD of posterior distribution by keyword "clinic name + regional name".

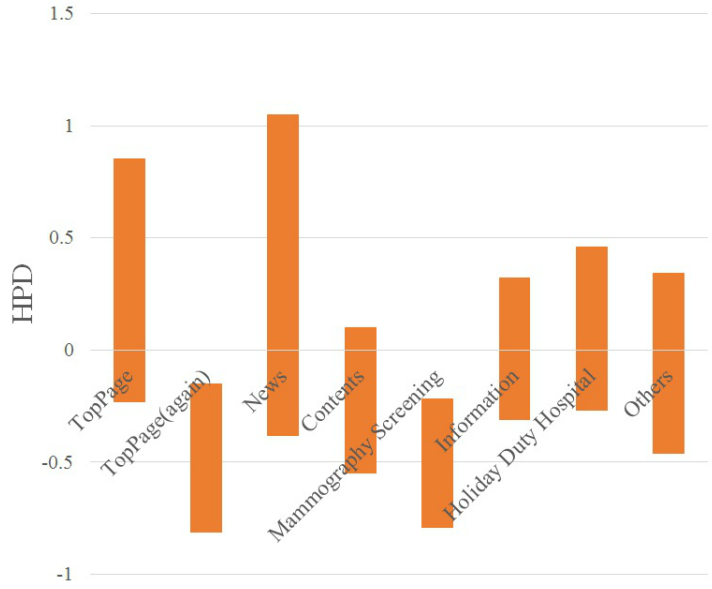
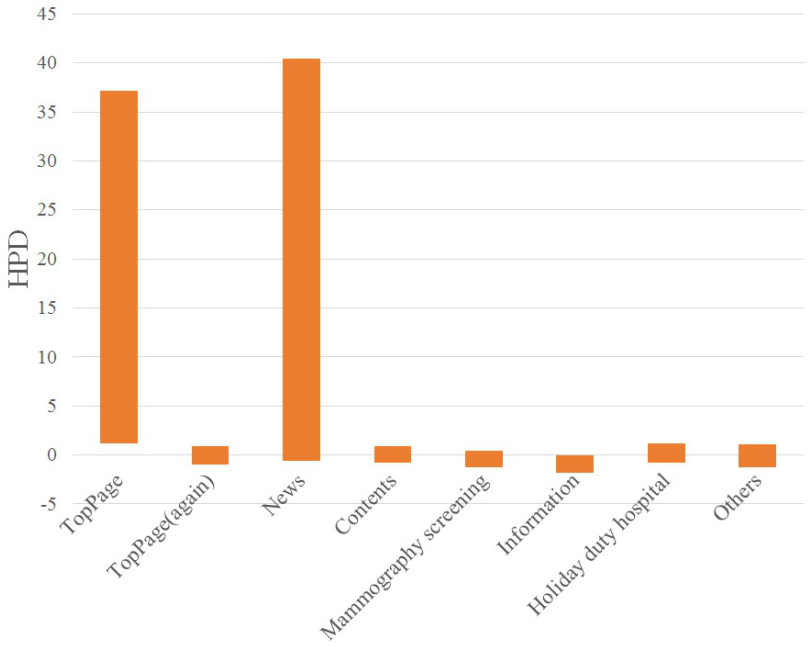



**Figure 8.** HPD of posterior distribution by keyword “clinic name + medical examination.”.






XSL•FO
RenderX

**Figure 9.** HPD of posterior distribution by keyword "mammography screening".

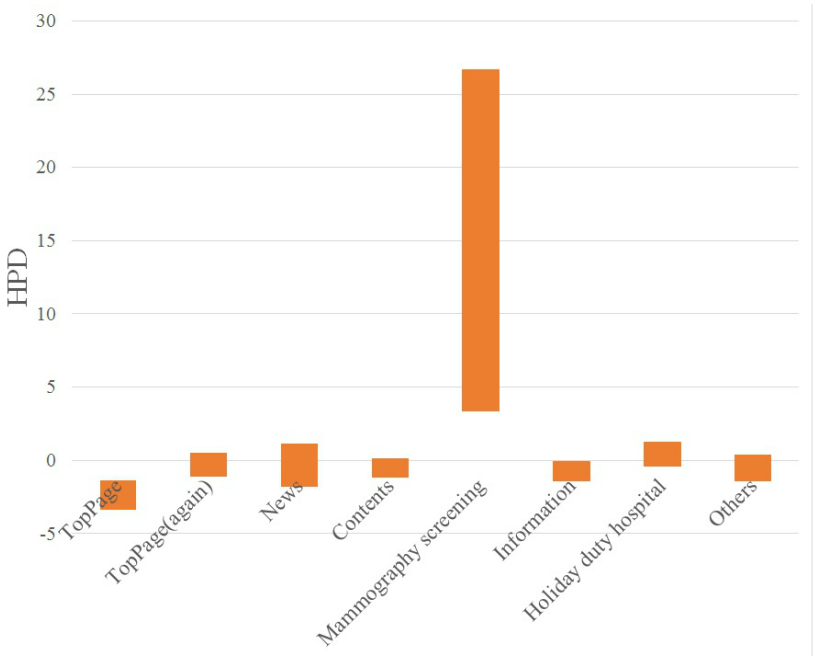



**Figure 10.** Posterior distribution by keyword "clinic name".

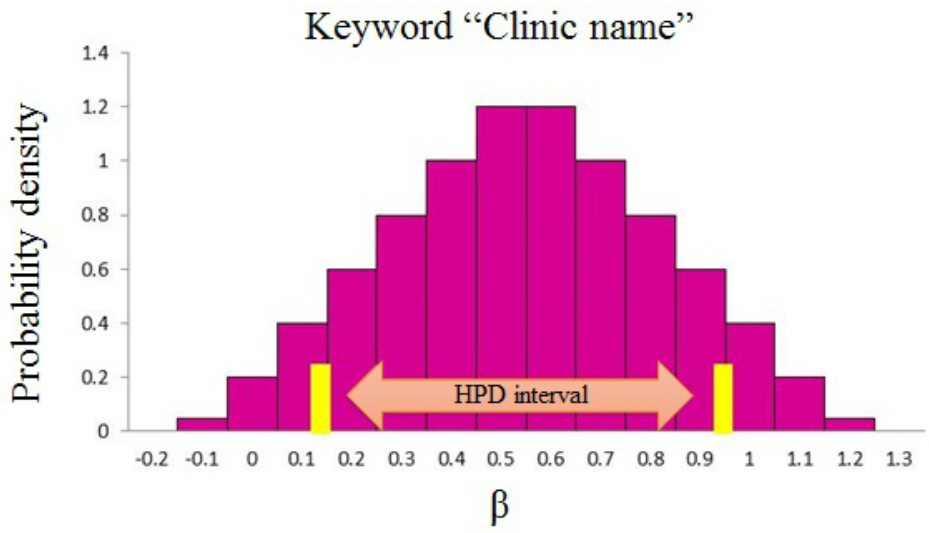
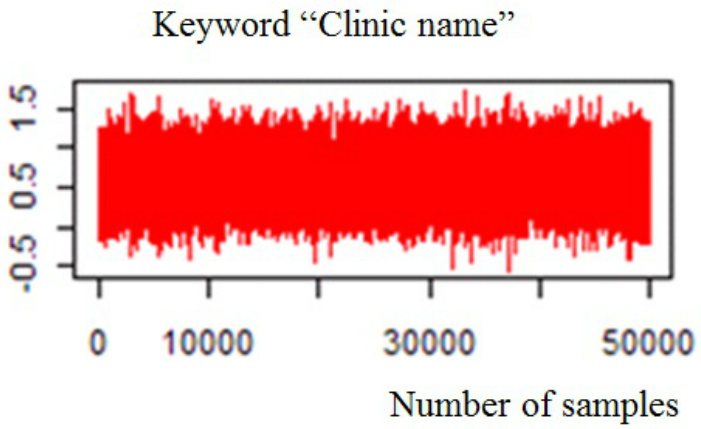


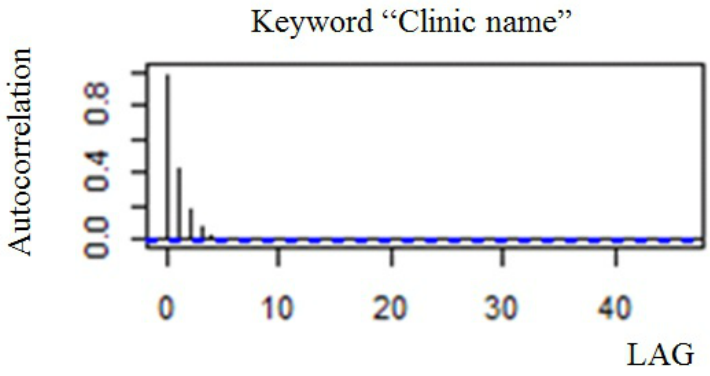

**Figure 11.** The simulation convergence situation by keyword "clinic name".






XSL•FO
RenderX

**Figure 12.** ACF by keyword "clinic name".




## Discussion

### Analysis of Searching Behavior of Medical Institution's Website Visitors

From the MCMC results, the ACF converged on most pages. This means that we obtained consistent results. Therefore, we expect our results to be generally valid.

When a visitor referred to the keyword “clinic name,” the visit probability to the main, website (again), and contents pages addressing medical examinations was positive. Thus, search by keyword “clinic name” had the effect of increasing the probability of visits to the main, website (again), and contents pages. In particular, it is possible that the primary concern of a visitor who referred to a keyword “clinic name” was to reach the contents page addressing medical examinations. The visit probability to the holiday duty hospital page was negative. Visitors to the holiday duty hospital information page did not refer to a clinic name, and it is possible that many people visited this page from other linked pages.

When a visitor referred to the keyword “clinic name and regional name,” the visit probability to the website (again) and the mammography screening page was negative. Search by keyword “clinic name and regional name” had the effect of decreasing the probability of visits to the website (again) and mammography screening pages. The visit probability to the website (again) was also low. The visitor using this keyword did not visit the website for a second time within the same session, so, it is likely that they were uninterested in the mammography screening page.

When visitors referred to the keyword “clinic name and medical examination,” the visit probability to the main page was positive, and the visit probability to the information page was negative. Search by keyword “clinic name and medical examination” had the effect of increasing the probability of visits to the main page and decreasing the probability of the visits to the mammography screening page. Some visitors who visited the contents of the medical examination in the keywords had a low probability of visiting the contents page of the medical examination. In such instances, we thought the visitor did not get the information they wanted. We concluded that the website did not lead its visitors to the page having the information they required. We suggest that the hospital administration changes its webpage design to address this problem.

When visitors referred to the keyword “mammography screening,” the visit probability to the mammography screening page was positive. Thus, the website did lead visitors who visited by the keyword “mammography screening” to the page they wanted. Search by keyword “mammography screening” had the effect of increasing the probability of visits to the mammography page. This indicated that the visitors could arrive at the page that they wanted. In this area, medical institutions that have implemented mammography screening are not many. Therefore, the results are expected.

As the visit probability to the main and information pages was negative, we concluded that visitors had no interest in these pages. Information about access to the clinic was published on the website. As the visit probability to the website was low, we concluded that visitors who referred to the keyword “mammography screening” had not yet become patients of the clinic.

These results reveal that the tendency of the visit probabilities in each category was different for different keywords. Therefore, it is possible to increase the visit probability to the page a visitor wants by understanding the search behaviors based on visitor needs and therefore improve website effectiveness.

### Problems and Overview

This study identified 4 problems for consideration.

### A Setup of an Interest Level

Page view, although used, could not be reflected in the result beyond recording the presence or absence of visits to each page, not the presence or absence of browsing behavior. How much browsing by visitors actually reflects their needs is unclear. Therefore, a model analysis that would include inspection time by visitors would improve our ability to gauge visitor interest.

### Six-Month Study Period

The fact that medical institution patient numbers fluctuate with the seasons should be taken into account. In this research, the access log covered a period of 6 months only; therefore, the fluctuations in patient numbers by season were not considered. In the study of medical institution websites, seasonality has not



XSL•FO
RenderX

been studied. As we believe that visitor access to website pages may also follow seasonal patterns, future research periods should also span one full year. Moreover, we think that a larger dataset is needed to accurately determine convergence in website access samples. Regardless of whether the amount of data used in this study was sufficient, we think that it is necessary to compare the analysis results using more data.

### Problem of a Prior Distribution Setup

In this study, because we had access only to data from the access log collection period, a noninformation prior distribution was assumed, so, no actual past data are reflected. To model a prior distribution, Ueda et al developed a prior distribution with high flexibility based on a nonparametric Bayesian model [15]. Bayesian estimation allows the determination of a posterior distribution, considering past data. We would like to consider using this technique for future research.

In this study, we used a noninformation prior distribution as the prior distribution. By setting the collected data with the prior distribution, we will be able to build a new model of medical institution webpage browsing behavior. In addition, using beta values obtained from improved prior distributions to estimate the behavior of website visitors, it becomes possible to build a website appropriate to medical institutions.

### Problem of Model Selection

Although the binomial probit model was used in this research, we could not verify the validity of the model. As the logit model, the classic Bayesian model, the nonparametric Bayesian model, and so forth are proposed in the literature as discrete selection models for use with a Bayesian approach, it is necessary to validate our model by comparing it with those of others [15-17]. One feature of medical institutions is that patient region and age groups differ by hospital scale, department, and region. In our research, because the only clinic website we targeted was the one near Sapporo, we could not address the characteristics of regionality or hospital scale. To better identify the browsing characteristics of visitors to the websites of many medical institutions, we would like to analyze other departments, regional areas, hospital scale, and so forth.

Recently, branding has become a marketing technique for hospital networks, and many patients select hospitals by recognizing their brands. In these situations, some hospitals are adopting a differentiated marketing strategy. Moreover, they are beginning to undertake customer relationship management (CRM), recognizing the lifetime value of a customer. As a hospital is an organization providing medical treatment as a service, it essentially has the same marketing challenges of any other company, while varying to a considerable degree in terms of the services provided for the public benefit. However, research by Kim states that the health care field can effectively apply CRM as well as any other field. The Bayesian approach used in this research is also a useful technique in CRM. As it is possible to perform heterogeneity modeling between consumers, this tool can be developed as one of the database marketing strategies for medical treatment [18].

For the modeling of heterogeneity among consumers, purchasing history data analysis, estimating heterogeneous price thresholds, and e-commerce site visitor behavior analysis has been conducted recently. However, behavior analysis on the websites of medical institutions remains unstudied [19-21].

Selection of medical institutions, as in the case of selecting products and services other than medical services, is affected by such competitive relationships between the patient's preferences and brand. In medical institution marketing activities, it is very important to know the variables. Therefore, it is necessary to further ascertain the heterogeneity between patients.

### Conclusion

To clarify the information that citizens want when searching the Web, we developed a searching behavior model for visitors to a medical institution's website using a Bayesian approach and determined the visit probability to each category of interest, classified by search keyword. We targeted the website access log of a clinic near Sapporo, for the January 1 to June 31, 2011 period and determined the visit probability to each category using the predetermined search keywords. In the case of the keyword "clinic name," the visit probability to the website, the website (again), and the contents page for medical examination was positive. For the holiday person-on-duty page, visit probability was negative. In the case of the keyword "clinic name and regional name," the visit probability to the website (again) and the mammography screening page was negative. In the case of the keyword "clinic name + medical examination," the visit probability to the website was positive, and the visit probability to the information page was negative. When visitors referred to the keywords "mammography screening," the visit probability to the mammography screening page was positive. Further analysis for not only the clinic website but also various other medical institution websites is necessary to build a general inspection model for medical institution websites; we want to consider this in future research. In addition, we hope to use the results obtained in this study as a prior distribution for future work and to conduct higher precision analysis.

## Acknowledgments

The authors gratefully acknowledge the work of past and present members of their laboratory. They would also like to thank the clinic and Dr. Katayama who provided data required for this study.

## Conflicts of Interest

None declared.



XSL•FO
RenderX

## Multimedia Appendix 1

R-code.

[PDF File (Adobe PDF File), 15KB - jmir_v18i7e199_app1.pdf ]

## References


1. Fukunaga W, Satomura Y. The environment and conditions for the establishment of Internet medicine. The Chiba Medical Society 2005;81:47-57.
2. The text of medical management talent. Japan: Ministry of Economy, Trade and Industry URL: http://www.meti.go.jp/report/whitepaper/ [accessed 2016-07-01] [WebCite Cache ID 6ifiMCyZ9]
3. Murakami T, Suyama A, Orihara R, TOSHIBA Corp. Research & Development Center. Consumer behavior modeling using Bayesian networks. 2006 Feb 11 Presented at: Conference of the Japanese Society for Artificial Intelligence; February 11, 2006; Japan. [doi: 10.11517/pjsai.JSAI04.0.260.0]
4. Saito Y. Behavioral Analysis of Website Visitor by Markov model. The operations research of Japan 2007:46-47.
5. Bucklin RE, Sismeiro C. A Model of Web Site Browsing Behavior Estimated on Clickstream Data. Journal of marketing research 2003;XL:249-267.
6. Şenel T, Cengiz MA. A Bayesian Approach for Evaluation of Determinants of Health System Efficiency Using Stochastic Frontier Analysis and Beta Regression. Comput Math Methods Med 2016;2016:2801081 [FREE Full text] [doi: 10.1155/2016/2801081] [Medline: 27118987]
7. Nobuhiko T. Marketing Analysis by Bayes Modeling. Tokyo, Japan: Tokyo Denki University Press; 2008.
8. Amano T, Nakazato H, Nakamura T. Web Mining Using Bayesian Method. The Institute of Electronics, Information and Communication Engineers 2004:43-48.
9. Google Analytics. URL: http://www.google.com [accessed 2016-07-01] [WebCite Cache ID 6ifos9Q39]
10. Yamamura M. Probit Analysis Based on ?National Livelihood Survey? Relevant to Use Nursing Care Service. Proceedings of Institute of Statistical Mathematics 2007 Jan 26:125-142.
11. Tsuchida N. A review on discrete choice models in marketing science. Journal of Business and Institutions 2010;8:63.
12. Satomura S. Data Science Learning Through R. Japan: Kyoritsu Shuppan Co., Ltd; 2010.
13. Jin M. Data Science by R. Japan: Morikita Publishing Co., Ltd; 2007.
14. Ogata Y. Biases and Uncertainties When Estimating the Hazard of the Next Nankai Earthquake. Journal of Geography 2001;110(4):602-614.
15. Markov Chain Monte Carlo Method. URL: http://www.omori.e.u-tokyo.ac.jp/ [accessed 2016-07-01] [WebCite Cache ID 6ifpRBT70]
16. Andrews RL, Ansari A, Currim IS. Hierarchical Bayes versus Finite mixture Conjoint Analysis models: A comparison of fit, prediction, and partworth recovery. Journal of Marketing Research 2002;39(1):87-98.
17. Ueda N, Yamada T. Introduction to Nonparametric Bayesian Models. Bulletin of the Japan Society for Industrial and Applied Mathematics 2007;17(3):196-214.
18. Kim Y, Shin DH. A Study on the Success Factor of Hospital CRM. Kawasaki Journal of Medical Welfare 2011;20:319-329.
19. Motohashi Y, Higuchi T. Shijyoukouzou no Henka wo kouryo shita Brand Sentaku ni yoru Koubairireki Data no Kaiseki. Japan Institute of Marketing Science (The Japanese Journal of Marketing Research) 2013;21:37-59.
20. Yamaguchi K. An Analysis of Website visit Behavior by a Hierarchical Bayesian Model Considering the Time Variation of Frequency. Japan Institute of Marketing Science (The Japanese Journal of Marketing Research) 2014;22:13-29.
21. Terui N, Wirawan D. Estimating Heterogeneous Price Thresholds. Marketing Science 2006;25:384-391.


## Abbreviations

**ACF:** autocorrelation function
**CRM:** customer relationship management
**MCMC:** Markov Chain Monte Carlo
**HPD:** highest posterior density



XSL•FO
RenderX

*Edited by G Eysenbach; submitted 16.09.15; peer-reviewed by H Potts, G Spyrou, E Nortey, NA Ismail; comments to author 29.03.16; revised version received 24.05.16; accepted 10.06.16; published 25.07.16.*

*Please cite as:*
*Suzuki T, Tani Y, Ogasawara K*
*Behavioral Analysis of Visitors to a Medical Institution's Website Using Markov Chain Monte Carlo Methods*
*J Med Internet Res 2016;18(7):e199*
*URL: http://www.jmir.org/2016/7/e199/*
*doi:10.2196/jmir.5139*
*PMID:27457537*

©Teppei Suzuki, Yuji Tani, Katsuhiko Ogasawara. Originally published in the Journal of Medical Internet Research (http://www.jmir.org), 25.07.2016. This is an open-access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/2.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on http://www.jmir.org/, as well as this copyright and license information must be included.



XSL•FO
RenderX

Original Paper

# First-Stage Development and Validation of a Web-Based Automated Dietary Modeling Tool: Using Constraint Optimization Techniques to Streamline Food Group and Macronutrient Focused Dietary Prescriptions for Clinical Trials

Yasmine Probst[1], BSc (Nutr), MSc (Nutr & Diet), MHlthInfo, GradCertBus, PhD; Evan Morrison[2], BCompSci, PhD; Emma Sullivan[1], BND (Hons); Hoa Khanh Dam[2], BCompSci, MAppSc, PhD

[1]School of Medicine, University of Wollongong, Wollongong, Australia

[2]School of Computing and Information Technology, University of Wollongong, Wollongong, Australia

**Corresponding Author:**
Yasmine Probst, BSc (Nutr), MSc (Nutr & Diet), MHlthInfo, GradCertBus, PhD
School of Medicine
University of Wollongong
Northfields Avenue
Wollongong, 2522
Australia
Phone: 61 42215302
Fax: 61 42214844
Email: yasmine@uow.edu.au

## Abstract

**Background:** Standardizing the background diet of participants during a dietary randomized controlled trial is vital to trial outcomes. For this process, dietary modeling based on food groups and their target servings is employed via a dietary prescription before an intervention, often using a manual process. Partial automation has employed the use of linear programming. Validity of the modeling approach is critical to allow trial outcomes to be translated to practice.

**Objective:** This paper describes the first-stage development of a tool to automatically perform dietary modeling using food group and macronutrient requirements as a test case. The Dietary Modeling Tool (DMT) was then compared with existing approaches to dietary modeling (manual and partially automated), which were previously available to dietitians working within a dietary intervention trial.

**Methods:** Constraint optimization techniques were implemented to determine whether nonlinear constraints are best suited to the development of the automated dietary modeling tool using food composition and food consumption data. Dietary models were produced and compared with a manual Microsoft Excel calculator, a partially automated Excel Solver approach, and the automated DMT that was developed.

**Results:** The web-based DMT was produced using nonlinear constraint optimization, incorporating estimated energy requirement calculations, nutrition guidance systems, and the flexibility to amend food group targets for individuals. Percentage differences between modeling tools revealed similar results for the macronutrients. Polyunsaturated fatty acids and monounsaturated fatty acids showed greater variation between tools (practically equating to a 2-teaspoon difference), although it was not considered clinically significant when the whole diet, as opposed to targeted nutrients or energy requirements, were being addressed.

**Conclusions:** Automated modeling tools can streamline the modeling process for dietary intervention trials ensuring consistency of the background diets, although appropriate constraints must be used in their development to achieve desired results. The DMT was found to be a valid automated tool producing similar results to tools with less automation. The results of this study suggest interchangeability of the modeling approaches used, although implementation should reflect the requirements of the dietary intervention trial in which it is used.

***(J Med Internet Res 2016;18(7):e190)*** doi:10.2196/jmir.5459



XSL•FO
RenderX

**KEYWORDS**

decision modeling; linear models; dietary requirements; clinical trial; food; programming, linear

## Introduction

The measurement of nutrients and prescription of foods for clinical studies can be a difficult task requiring consideration of a number of different elements [1]. When conducting a dietary intervention, it is imperative that researchers ensure consistent allocations of macro- and micronutrients across participants, while also tailoring the required dietary approach to the food requirements, for example, food preferences, of the individual participants as needed. Many high-quality published dietary studies do not report on the use of such tailored approaches and rather apply the underlying assumption that usual dietary intake will be maintained by participants throughout the duration of a trial [2]. However, within a randomized controlled trial design, intervening using a food-/nutrient-based approach will inherently result in changes to dietary intake during the trial, potentially affecting the outcomes. This is particularly evident when a target food is provided to participants, with studies showing that the target food will be eaten in addition to rather than substituted into the usual diet [3], resulting in increased energy (calories or kilojoules) intake due to the intervention. To reduce the effect of these changes, predefined, personalized, energy-focused dietary prescriptions should be implemented via the use of dietary modeling before the intervention.

Modeling is employed to test logic, demonstrate a concept or an idea, and serve as a representation of reality. It often has a mathematical basis [4]. The term "model" also implies variability of the outcomes; hence, multiple options are often tested. In practice, modeling is a theoretical process allowing different scenarios [5] to be created. These scenarios typically stem from an idea, concept, or change of practice. For dietary modeling, the concept generally relates to achieving dietary targets (food or nutrient) by consumption of given amounts (servings) from a range of food groups. It utilizes a combination of food consumption data and/or food composition data [5], although it is not limited to these. In one example of modeling of dietary intake, the concentration of a known nutrient *within* a food is multiplied by the amount of the food consumed to determine the contribution to the total nutrient intake. This type of modeling requires access to food consumption data, such as that of a national survey [6,7]. Modeling conducted by the Australian regulatory authority, Food Standards Australia New Zealand, use tailored software based on the SAS statistical package using the following equation: *dietary intake=Σ(nutrient/chemical concentration × food consumption)* [6]. The software was custom developed to create high-level dietary models at a population level to address chemical exposure and the effect of food contamination [8]. Modeling may also be used to standardize dietary intake across a participant group at the point of intervention such as in a dietary intervention trial, although resource limitations are likely to reduce the opportunity to develop customized software.

The use of tailored tools, specifically for dietary modeling, has the potential to aid the translation from nutrient to food information and incorporate nutrition recommendations [4] related to diet-disease relationships. The tools should ensure a consistent and streamlined process is applied across the entire trial to minimize variability. Manual approaches to prescribing individualized diets are common to dietary studies [9-13]. They are heavily user dependent, providing the potential for a high degree of variability between prescribed outcomes. Manual methods do, however, allow some consideration of the practical issues related to dietary prescriptions at an individual level and food-based guidelines to be incorporated. Practice-based examples include a Microsoft Excel spreadsheet that is manually manipulated to achieve food group targets [14,15] whereby the user, often a dietitian or nutritionist, aims to achieve energy (calories or kilojoules) and macronutrient (total fat, protein, carbohydrate) proportions with minimal variability from the overall trial targets. Modeling is achieved by manually adjusting the number of prescribed servings within a given range (commonly based on dietary guidelines). This process is cumbersome and time consuming for the user and may have significant resource implications within a clinical trial if the dietitian is also needed to monitor the dietary intake of the participants.

Alternatively, adding automation to the modeling process has the potential to decrease the variability and time taken to create the models while still employing a user-dependent approach. The approach utilized to formulate the Australian Dietary Guidelines is an example of modeling with an element of automation, using the Solver Platform for Microsoft Excel [16]. Solver applies a computerized method for finding optimal solutions using predetermined constraints (data limits) to Excel-based spreadsheets. These constraints must be imputed into the tool before it is applied to a spreadsheet. Solver has the potential to minimize the user burden by minimizing inconsistencies from trial targets at a group level but does require extensive input of constraint details. Additionally, Solver does not present the ability to easily optimize across competing objectives that may be found in a clinical trial as it has been designed to primarily provide user-defined constraint satisfaction. The aforementioned spreadsheet-based tools are also limited to food-based output only. They do not consider the need for individualized modeling, a process that would require separate calculation of estimated energy requirements (EER). This individualized approach is common to the highly controlled environment of a dietary intervention clinical trial where each individual needs a targeted dietary prescription rather than a generic one.

The process of dietary modeling in food-based clinical trials when performed manually required dedicated time and resource commitments and the need to calculate energy requirements before development of each model. By applying constraint optimization techniques to this process, it can be automated, saving both time and resources and streamlining the overall approach used. The development of the Dietary Modeling Tool (DMT) [17] has the potential to provide an automated method for dietary prescription, tailored to individual characteristics within the food and/or nutrient-based constraints of a clinical



XSL•FO
RenderX

trial. The objectives of the DMT were that it would take on a simple, Web-based (widely accessible), and user-friendly format and target the individual energy requirements of participants. Its development would result in a reduction of between user variability that may become evident when calculating target servings for food groups based on energy requirements using manual and partially automated approaches. Development of streamlined models would also allow users who are not trained in dietetics to create the models, and the dietitians would be employed to address only those models where specific considerations such as food preferences or avoidances need to be addressed. The context for the DMT applied the following assumptions. The clinical trial (study) targets would be generated based on selected food groups to match the nutrition targets defined in the trial. Modeling targets may be specific to the trial or may relate to default nutrition guidelines such as the Dietary Guidelines and Nutrient Reference Values. The clinical trial would have baseline measures for all participants, that is, age, gender, height, and weight (also used to calculate body mass index), available to generate individualized models. This type of demographic information is commonly collected. The overall aim of the tool was that the developed DMT could be easily reconfigured to use multiple macronutrient trial targets; could be adjusted to accommodate a wide range of participant dietary preferences, for example, vegetarian dietary patterns; and could be used across multiple studies, that is, maintain the default preferences for studies A, B, C...X, which may be occurring concurrently. This paper describes the first-stage developmental process of a tool to automatically perform dietary modeling using food group and macronutrient requirements as a test case. The DMT was then compared with existing approaches to dietary modeling (manual and partially automated), which were previously used by dietitians for dietary modeling.

## Methods

The development of the DMT applied lessons from an existing tool developed for the Australian food guidance system (AFGS) that used a *linear* programming approach to modeling. Algorithms published as part of the AFGS were used as the basis for developing the DMT using *nonlinear* modeling. Constraint optimization was also used to ensure the DMT was suited to developing individual dietary prescriptions that are needed in dietary intervention trials as the AFGS targeted population groups. The developed algorithms were combined with existing manual advice models [18] that had been used in published dietary intervention trials conducted by the Smart Foods Centre, University of Wollongong. Second, this paper describes the comparative validation of the DMT against an existing validated manual modeling tool and compares the output with partially automated approaches to dietary modeling. It was hypothesized that the DMT would provide clinically valid results when compared with its manual counterparts while also minimizing user burden.

### The Constraint Optimization Problem

Formally, a constraint is a function $CF(Dx_i,...,Dx_j)$ resulting in a Boolean output {True, False}. The result is true if the combination of all values is allowed and false otherwise. An objective function $OF(Dx_i,...,Dx_j)$ results in a set of Solutions R, where R is a set of solutions that are optimal with respect to the objectives OF. A constraint optimization problem is a tuple (X,D,C,O) where X is a set {x1,...,xm} of variables, D is a set {d1,...,dm} of variable domains, C is a set {c1,...,cj} of constraints defined over X, and O is a set {o1,...,op} of objective functions defined over X [19].

### Food Guidance Modeling (Using Linear Programming)

Modeling conducted for the AFGS [20] used linear programming [21,22] with population data. Linear programming was used to allow more complex diets to be created within energy and macronutrient restrictions. For example, if a food item or a food group referred to as *i (i=1...m)* is consumed in set amounts shown as $s_i$ grams (ie, the target serving size), then the daily total intake of a given nutrient *j* is shown by the summation $\Sigma\ n_i.s_i.(c_{ij}/100)$, where $n_i$ = the consumed number of servings of the food item or food group *i* and $c_{ij}$ is the amount of the nutrient *j* per 100 g of the food item or food group *i*. If 100 g of the food item or food group *i* provides $e_i$ kilojoules of energy to the daily diet, the dietitian then needs to decide on the set of servings $(n_i \geq 0, i=1...m)$ from a total list of all possible foods items or food groups *m*, such that the daily total energy intake of *m* is

$$E=\Sigma\ n.s.e \quad [1]$$

and *i=1* is minimized, and such that the guidelines for each of the "nutrients" *j=1...v* are all simultaneously satisfied, that is,

$$m\Sigma\ ni.si.cij/100^3\ NRVj,\ j=1...v$$

Using this approach, it may be necessary to break down the information by food groupings to achieve an outcome.

Further consideration then needs to be given to cultural norms, food preferences, intolerances, palatability, and a number of other factors by applying fixed, a "no more than" or "no less than," constraints to particular elements of the equation. For example, a maximum intake level might be applied to a food group. At the population level, individual considerations such as food allergies are not addressed. When limitations are applied to the aforementioned process with the aim to minimize energy intake, in order to minimize population obesity levels, a number of solutions may be apparent. These solutions (expressed as numbers) are not always whole integers. As food is generally consumed as a whole item, rounding is required to allow recommendations to be made for a given frequency of consumption, for example, per day or per week. This process results in some minor adjustment to the target energy intake, although this adjustment was found to be negligible [20].

### Dietary Modeling Tool Development

Constraint optimization techniques were used to determine whether nonlinear constraints are better suited to the modeling process when a smaller sample (such as the participants of a dietary randomized controlled trial) is targeted. Constraint optimization is a process whereby a set of rules is created and refined and needs to be upheld for a model to be developed. A constraint optimization problem is a problem that is used to find an optimal assignment of values to a given a set of variables, their domain, a constraint function, and an objective function.



XSL•FO
RenderX

A set of variables x1,...,x2 represent values that can be changed; their domain are the acceptable values that each variable can be assigned. For example, given a variable for "servings of vegetables per day" an acceptable domain is 0 to 10. A constraint over the set of variables is a restriction on variables; for example, the energy (calories or kilojoules) of all servings must be less than X. A constraint function is an abstract function that takes a set of variables and their values and returns a True or False answer indicating if they violate the constraints or not. True is returned if all of the constraints are satisfied. A constraint solver is an application that takes in variables and their domain and changes the values of the variables until it can find a satisfactory solution. A satisfactory solution is one where all of the constraints are satisfied and the constraint function returns true. An objective function is a function that takes a set of solutions and determines the best given measure; for example, the objective function "minimize kilojoules" will take a set of results found by the constraint solver and pick the result with the lowest number of kilojoules.

In contrast to the AFGS, it is likely that these constraints will follow a nonlinear form because of the varied considerations needed when modeling a diet, as outlined earlier. The DMT will draw on existing food consumption data collected from published dietary intervention trials as the basis for the weighting of food groups rather than population-focused food consumption data as was applied to AFGS. Output will be provided as the number of servings of key food groups required by a participant, of given energy requirements, to meet the criteria of the clinical trial in which it is being used, optimized against key food group serving suggestions. This is so that a solution diet is not composed of one single food group or presented based on an irregular food group split.

Development of the DMT followed a stepwise process. Food data from completed dietary trials [23-25] were pooled and the percentage contribution of common food groups determined based on macronutrient composition. These food groups were rank ordered under each macronutrient (total fat, protein, carbohydrate) to determine the primary sources. Foods seen to contribute to <75% of the total for each macronutrient were taken as the top foods consumed. Individual foods belonging to the food groupings determined were then categorized by the relative proportions of all macronutrients contained in a single serving. From this the mean energy (kilojoules) and macronutrient content (grams) were determined. Subcategories of food groupings were created based on secondary macronutrients (saturated, monounsaturated, and polyunsaturated fatty acids) and on other nutritive components such as the presence of starch or sugar, which may be required for particular participant groups such as persons with diabetes.

For each food group, the mean nutrient content, standard deviation (SD), range, and coefficient of variation (*CV=SD/mean × 100*) were determined for all foods within that group. Acceptable variation was set at a CV of < 15% for the macronutrients; otherwise, acceptability of the variation was assessed by comparison with an existing food guidance system. Standard deviation and range results were compared with those reported in the 2003 American Diabetes Association Exchange Lists for Meal Planning lists [26]. These lists were considered to be the only comprehensive food exchange lists suited to provide specific data on within-list variations from mean nutrient estimates.

### *Setting Up the Interface*

Developed online, the DMT [17] relies on 2 data sources that are not seen by the user to populate the nutrient data (eg, macronutrients) for each of the food groups (eg, vegetables, fruits, and so on):

1. NUTTAB, a reference food composition database [27] for Australia containing a list of all available foods, food groups, their energy, and macro- and micronutrient composition.

2. Dietary intervention trial database containing pooled baseline food intake data from completed trials before intervention.

An overview of the process is shown in Figure 1. In summary, to use the tool, users (likely dietitians) access a website [17]. Initially, the user provides as input to the tool the following data related to the trial:

1. Macronutrient targets for the trial including total fat, protein, and carbohydrate.

2. Target servings *T1, T2, T3,..., Tn* where *Ti* is the target serving for a food group. For example, T1 is the target serving for vegetables (eg, 5 serves), T2 for grains (eg, 6.5), T3 for fruits (eg, 3).

These values are maintained across the trial, and adjustment will be applied to all models created for participants of that trial. Via a separate interface (see Multimedia Appendix 1), the user then inputs the participant details (including height, weight, age, and gender).

The tool then computes the EER from resting energy expenditure (REE) [28] using the following formulae:

For females, REE=9.99×weight+6.25×height–4.92×age–161

For males, REE=9.99×weight+6.25×height–4.92×age+5

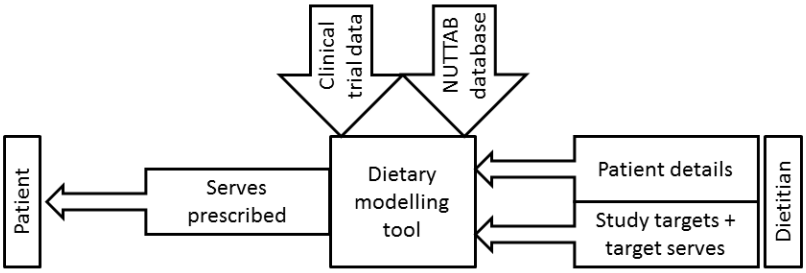
EER=REE×PA

Physical activity (PA) is accounted for by standardized activity factors as used in dietetic practice. For the purpose of this first-stage development an activity factor of 1.6 (light activity) was applied [29]. The user may then enter any study-specific macronutrient percentages and desired food group servings (lean meats, dairy, and so on) to suit the participant food preferences. For example, if the participant follows a vegetarian diet the food group servings for meat may be removed by the dietitian and replaced with meat alternatives. This would return to the default trial criteria for the next participant who may not follow a vegetarian diet. The automated DMT would then provide the user with target servings per food group to meet the trial requirements suited to each participant.



XSL•FO
RenderX

**Figure 1.** A schematic of the underlying process of using the Dietary Modeling Tool. Note: a dietitian defines the patient details and study targets and target servings per food group and/or per nutrient (study specific), which is entered into the Dietary Modeling Tool.The tool draws data from both a clinical trial and NUTTAB (nutrient tables) database to provide information related to the number of servings suited to the patient details (gender, height, weight, age). These servings are provided to the patient for implementation of the dietary approach.




### *Creation of the Models*

Then, let *X1*, *X2*,..., *Xn* be the servings described for the food groups, for example, X1 for vegetables and X2 for grains. The DMT is needed to determine the "best" value for Xi that meet the constraints and objective functions. Therefore, for carbohydrate (CHO), let TotalCHO be the total of carbohydrates in the servings prescribed shown in Figure 2, equation (a), where *CHOi* is the CHO nutrient data for *Xi*. Similarly, total protein and fat as per equations (b) and (c) in Figure 2 respectively. The resultant total energy of the servings prescribed is then calculated as:

**Figure 2.** Constraint and objective functions for calculation of food groups based on carbohydrate, protein and fat content.

$$(a)\ Total\ CHO = \sum_{i=1}^{n} \mathrm{Xi} * \mathrm{CHOi}$$
$$(b)\ TotalPTN = \sum_{i=1}^{n} \mathrm{Xi} * \mathrm{PTNi}$$
$$(c)\ TotalFAT = \sum_{i=1}^{n} \mathrm{Xi} * \mathrm{FATi}$$

TotalEnergy = TotalCHO + TotalPTN + TotalFAT,

with the percentage energy for macronutrients calculated as given below.

%CHO = [(TotalCHO × 17)/TotalEnergy] × 100

%PTN = [(TotalCHO × 17)/TotalEnergy] × 100

%FAT = [(TotalCHO × 36)/TotalEnergy] × 100

The Euclidean distance between (%CHO, %PTN, and %FAT) and (TargetCHO, TargetPTN, and TargetFAT), referred to as *d(CHO, PTN, FAT)*, was therefore calculated as seen in Figure 3, and, in turn, the tool needs to find the servings prescribed, *Xi*, such that *d(CHO, PTN, FAT)* is minimized, that is, as close to 0 as possible, while the following constraints are satisfied:

1. For every food group i, Xi ≥ Ti (ie, the servings prescribed are greater than or equal to the target serving for every food group).

2. TotalEnergy ≤ EER (ie, the total energy of the servings prescribed is less than or equal to EER).

By example, using the following constraints for prescribed servings the following might be used: $X_1 \geq 6.5$/day (grains), $X_2 \geq 3$/day (fruits), $X_3 \geq 3$/day (dairy).

**Figure 3.** Equation for calculating the Euclidean distance between carbohydrate (CHO), protein (PTN) and fat.

$$d(CHO, PTN, FAT) = \sqrt{(\%CHO - TargetCHO)^2 + (\%PTN - TargetPTN)^2 + (\%FAT - TargetFAT)^2}$$

## Validation and Use in Practice

To evaluate validity, the automated process was compared with a partially automated and manual diet modeling process, and the consistency in servings was compared using the different methods, for a variety of energy levels. As there is no gold standard for validation of dietary models, the manual approach was considered the standard of reference.

For the manual process, a Microsoft Excel calculator [18] used in previous clinical trials conducted at the University of Wollongong was used, and for the partially automated process Microsoft Excel Solver [16] was applied. As all tools were developed based on the same underlying process of food groupings, a consistent food group–based approach could be compared across tools (for vegetables, breads and cereals, fruits, low fat dairy, lean meat, cheese, eggs, oily fish, monounsaturated fatty acid products, and polyunsaturated fatty acid products). Models were created in accordance with the recommendations of the Australian Guide to Healthy Eating [30]. They provided outcomes for the number and size of servings for each food group within the model. Diet models, by food group, were created for 5000 kJ through to 10000 kJ daily intakes, in 500-kJ increments. For comparative purposes the percentage of energy provided from macronutrients was set to be 50% carbohydrate, 20% protein, and 30% fat for all approaches (manual, partially automated, and DMT).

To create comparative data, EER values using sample participant height, weight, age, and gender were calculated using a separate spreadsheet. These data were applied to the Excel calculator and Solver modeling tools. Data for the number of food groups



XSL•FO
RenderX

and calculated energy and macronutrient levels were provided by each of the tools.

### *Excel Spreadsheet Calculator (Manual)*

An experienced Accredited Practising Dietitian (APD) manually manipulated the number of target servings of food groups in order to achieve appropriate diet models for each kilojoule increment. For repeatability, a less experienced dietitian also performed the same task separately using the Excel spreadsheet calculator [18]. This process created a second comparative set of dietary models for each of the kilojoule targets.

### *Excel Solver (Partially Automated)*

The Excel Solver add-in was applied to the aforementioned calculator. In order to determine the appropriate output and ensure consistency, the constraints outlined in Table 1 were applied.

### *Dietary Modeling Tool (Automated)*

The sample data for a participant's height, weight, age, and gender were entered *directly* into the DMT. Trial targets in the study interface were set to the default macronutrient distributions as outlined above.

**Table 1.** Number of serving constraint details per food group applied to the Microsoft Excel Solver modeling tool based on the study by Gillen and Tapsell [19,26].

| Food group used for modeling | Number of serving constraint details (per day) | Additional number of serving constraints (required for 8500-10,000 kJ models) |
|---|---|---|
| Vegetables | ≥ 5, [a] | |
| Whole grains | ≥ 4, | |
| Fruits | ≥ 2, <4 | |
| Sugar | ≤ 3 | Unrestricted |
| Milk/yoghurt (low/reduced fat) | ≥ 2.5 | |
| Milk/yoghurt (whole) | ≤ 0 | |
| Soy milk (whole) | ≤ 0 | |
| Meat (lean choice, per 30 g) | ≥ 3 | ≥ 5 |
| Cheese (reduced fat, per 30 g) | ≥ 0 | |
| Eggs (per 30 g) | ≥ 0 | |
| Oily fish | ≥ 0.43[b] | |
| Monounsaturated fatty acids | ≥ 0 | |
| Polyunsaturated fatty acids | ≥ 0 | |

[a] : No upper constraint limit.

[b] Equates to at least 1 serving per week.

### Data Analysis

The variability of the percentage of energy from each of the macronutrient targets was calculated for all tools. The outcomes from each of the dietary modeling approaches were compared for grouped food data for each of the kilojoule increments tested. The percentage difference of each of the methods, in comparison with the reference method (output from the manual process created by an experienced dietitian), was calculated in order to determine the comparative validity of the processes.

## Results

Constraint optimization was found to be a suitable approach to tool development. As outlined earlier, the DMT was developed with 2 interfaces for users: a study interface, for defining default constraints for a study, and a user interface for modeling individual participant diets within the selected trial. Screenshots of these user screens are provided in Multimedia Appendix 1 [17]. This approach to development allows multiple users to model diets in different studies simultaneously. The forms were hosted online [17] also allowing multi-user access from varied locations.



XSL•FO
RenderX

**Table 2.** Target servings prescribed for each food group using the manual, partially automated, and automated approaches applied to different energy frameworks.

| No. | Model[a] (kJ[b] target) | Vegetables (%Δ) | Grains (%Δ) | Fruits (%Δ) | Dairy (%Δ) | Lean meat, 30 g (%Δ) | Cheese[c], 30 g (%Δ) | Eggs[d], 1 egg (%Δ) | Fish[e], 90 g (%Δ) |
|---|---|---|---|---|---|---|---|---|---|
| 1. | Reference (5000) | 5.00 | 5.00 | 2.00 | 2.00 | 3.00 | 0.29 | 0.58 | 0.43 |
| | Partially automated | 5.00 (0) | 4.00 (20) | 2.00 (0) | 2.50 (25) | 3.00 (0) | 0.00 (100) | 0.00 (100) | 0.43 (0) |
| | DMT[f] | 5.00 (0) | 4.50 (10) | 2.00 (0) | 2.00 (0.) | 3.00 (0) | 0.00 (100) | 0.00 (100) | 0.00 (100) |
| | Manual | 5.00 (0) | 5.00 (0) | 2.00 (0) | 2.50 (25) | 3.00 (0) | 0.00 (100) | 0.29 (51) | 0.43 (0) |
| 2. | Reference (5500) | 5.00 | 5.00 | 2.00 | 2.00 | 3.50 | 0.29 | 0.58 | 0.43 |
| | Partially automated | 5.00 (0) | 4.00 (20) | 2.30 (15) | 2.50 (25) | 3.00 (14) | 0.00 (100) | 0.00 (100) | 0.43 (0) |
| | DMT | 6.00 (20) | 7.00 (40) | 2.07 (4) | 2.00 (0) | 3.00 (14) | 0.00 (100) | 0.00 (100) | 0.00 (100) |
| | Manual | 5.00 (0) | 5.00 (0) | 2.00 (0) | 2.50 (25) | 3.00 (14) | 0.00 (100) | 0.29 (51) | 0.43 (0) |
| 3. | Reference (6000) | 5.00 | 6.00 | 2.00 | 2.00 | 3.50 | 0.29 | 0.58 | 0.43 |
| | Partially automated | 5.00 (0) | 4.00 (33) | 2.99 (50.) | 2.50 (25) | 3.00 (14) | 0.00 (100) | 0.00 (100) | 0.43 (0) |
| | DMT | 6.44 (29) | 7.00 (17) | 2.50 (25) | 2.17 (9) | 3.00 (14) | 0.00 (100) | 0.00 (100) | 0.12 (73) |
| | Manual | 5.00 (0) | 6.00 (0) | 2.00 (0) | 2.50 (25) | 3.50 (0) | 0.00 (100) | 0.29 (51) | 0.43 (0) |
| 4. | Reference (6500) | 5.00 | 6.00 | 2.00 | 3.00 | 4.00 | 0.29 | 0.58 | 0.43 |
| | Partially automated | 5.00 (0) | 4.21 (30) | 4.00 (100) | 2.50 (17) | 3.00 (25) | 0.00 (100) | 0.00 (100) | 0.43 (0) |
| | DMT | 7.00 (40) | 7.38 (23) | 2.50 (25) | 2.49 (17) | 3.20 (20) | 0.09 (69) | 0.03 (96) | 0.43 (0) |
| | Manual | 5.00 (0) | 6.00 (0) | 2.00 (0) | 2.50 (17) | 3.50 (13) | 0.29 (0) | 0.29 (51) | 0.43 (0) |
| 5. | Reference (7000) | 5.00 | 6.00 | 2.00 | 3.00 | 4.00 | 0.29 | 0.58 | 0.43 |
| | Partially automated | 5.00 (0) | 5.21 (13) | 4.00 (100) | 2.53 (16) | 3.31 (17) | 0.09 (69) | 0.16 (72) | 0.18 (58) |
| | DMT | 7.00 (40) | 9.25 (54) | 2.50 (25) | 2.51 (16) | 3.00 (25) | 0.04 (85) | 0.00 (99) | 0.17 (60) |
| | Manual | 5.00 (0) | 6.00 (0) | 3.00 (50) | 2.50 (17) | 3.50 (13) | 0.29 (0) | 0.29 (51) | 0.43 (0) |
| 6. | Reference (7500) | 5.00 | 7.00 | 2.00 | 3.00 | 4.00 | 0.29 | 0.58 | 0.86 |
| | Partially automated | 5.00 (0) | 5.78 (17) | 4.00 (100) | 3.00 (0) | 3.04 (24) | 0.00 (100) | 0.00 (100) | 0.43 (50) |
| | DMT | 7.00 (40) | 9.25 (32) | 4.00 (100) | 3.07 (2) | 3.00 (25) | 0.00 (100) | 0.01 (99) | 0.86 (0) |
| | Manual | 5.00 (0) | 7.00 (0) | 3.00 (50) | 2.50 (17) | 4.00 (0) | 0.00 (100) | 0.29 (51) | 0.43 (50) |
| 7. | Reference (8000) | 5.00 | 8.00 | 3.00 | 3.00 | 4.50 | 0.29 | 0.86 | 0.86 |
| | Partially automated | 5.00 (0) | 6.27 (22) | 4.00 (33) | 3.00 (0) | 4.50 (0) | 0.00 (100) | 0.00 (100) | 0.43 (50) |
| | DMT | 7.00 (40) | 9.25 (16) | 4.00 (33) | 4.00 (3) | 3.00 (33) | 0.26 (10) | 0.00 (100) | 0.75 (13) |
| | Manual | 5.00 (0) | 7.00 (13) | 3.00 (0) | 2.50 (17) | 5.00 (11) | 0.29 (0) | 0.58 (33) | 0.43 (50) |
| 8. | Reference (8500) | 5.00 | 8.00 | 3.00 | 3.00 | 5.00 | 0.43 | 0.86 | 0.86 |
| | Partially automated | 5.00 (0) | 7.34 (8) | 4.00 (33) | 3.00 (0) | 5.00 (0) | 0.00 (100) | 0.00 (100) | 0.43 (50) |
| | DMT | 7.00 (40) | 9.25 (16) | 4.00 (33) | 4.00 (33) | 3.32 (34) | 0.26 (40) | 0.00 (100) | 0.43 (50) |
| | Manual | 5.00 (0) | 7.50 (6) | 3.00 (0) | 3.00 (0) | 5.00 (0) | 0.29 (33) | 0.58 (33) | 0.43 (50) |
| 9. | Reference (9000) | 5.00 | 9.00 | 3.00 | 3.00 | 5.00 | 0.29 | 0.86 | 0.86 |
| | Partially automated | 5.00 (0) | 8.12 (10) | 4.00 (33) | 3.00 (0) | 5.00 (0) | 0.00 (100) | 0.04 (95) | 0.58 (33) |
| | DMT | 7.00 (40) | 9.25 (3) | 4.00 (33) | 4.00 (33) | 5.00 (0) | 0.29 (0) | 0.17 (81) | 0.50 (42) |
| | Manual | 5.00 (0) | 8.00 (11) | 3.50 (17) | 3.00 (0) | 5.00 (0) | 0.43 (50) | 0.58 (33) | 0.86 (0) |
| 10. | Reference (9500) | 5.00 | 9.00 | 4.00 | 3.00 | 6.00 | 0.43 | 0.86 | 0.86 |
| | Partially automated | 5.00 (0) | 9.01 (0) | 4.00 (0) | 3.00 (0) | 5.23 (13) | 0.00 (100) | 0.18 (79) | 0.58 (33) |
| | DMT | 7.00 (40) | 9.25 (3) | 4.00 (0) | 4.00 (33) | 5.00 (17) | 0.29 (33) | 0.57 (33) | 0.50 (42) |
| | Manual | 5.00 (0) | 8.50 (6) | 4.00 (0) | 3.00 (0) | 5.00 (17) | 0.43 (0) | 0.58 (32) | 0.86 (0) |



XSL•FO
RenderX

| No. | Model[a] (kJ[b] target) | Vegetables (%Δ) | Grains (%Δ) | Fruits (%Δ) | Dairy (%Δ) | Lean meat, 30 g (%Δ) | Cheese[c], 30 g (%Δ) | Eggs[d], 1 egg (%Δ) | Fish[e], 90 g (%Δ) |
|---|---|---|---|---|---|---|---|---|---|
| 11. | Reference (10,000) | 5.00 | 9.00 | 4.00 | 3.00 | 6.00 | 0.43 | 1.43 | 0.86 |
| | Partially automated | 5.00 (0) | 10.07 (12) | 4.00 (0) | 3.00 (0) | 5.45 (9) | 0.05 (88) | 0.15 (90) | 0.55 (36) |
| | DMT | 7.00 (40) | 9.25 (3) | 4.00 (0) | 4.00 (33) | 5.00 (17) | 0.29 (33) | 0.57 (60) | 0.50 (42) |
| | Manual | 5.00 (0) | 8.50 (6) | 4.00 (0) | 3.00 (0) | 6.00 (0) | 0.43 (0) | 0.86 (40) | 0.86 (0) |

[a] Reference method employed was the use of a manual spreadsheet-based tool used by an Accredited Practising Dietitian. Partially automated process applied Microsoft Excel Solver application.

[b] kJ: kilojoule.

[c] 0.14=1/week, 0.286=2/week.

[d] 0.286=1/week, 0.58=2/week, 0.86=3/week.

[e] 0.43=1/week, 0.86=2/week.

[f] DMT: Dietary Modeling Tool.

When data were compared with the other modeling forms, the general trend for the output across the 4 dietary models in each kilojoule increment was relatively similar (Table 2). The greatest consistency was seen between the 2 manual approaches. Notably, for a considerable number of food groups across the kilojoule targets, there was no difference between these 2 models, justifying use of the manual process as the standard of reference due to repeatability of the data. The DMT outcomes were the most varied from those created using the reference process. The outcomes for monounsaturated fatty acids and polyunsaturated fatty acids (subnutrients) across the modeling tools were highly variable, with differences of up to 200% for the 5000-kJ and 5500-kJ targets, data not shown. In this instance the reference model was prescribing 1 teaspoon monounsaturated fatty acids, as opposed to 3 teaspoons from the DMT. Less variation was evident for these particular food groups in the higher kilojoule targets (≥8000 kJ). The lean meat prescription was largely consistent across all of the dietary models. The largest difference in the lean meat outcomes was seen for the 8000-kJ and 8500-kJ targets, with 33% and 34% differences, respectively, between the DMT and reference model. This difference equated to one and a half servings (one serving = 30 g) of lean meats. Furthermore, the prescription of fish from the 5000-kJ to the 7000-kJ targets for the reference method and solver models all equated to 90-g servings of fish per week. The DMT produced the same results for the 6500-kJ target; however, for kilojoule targets between 5000 kJ and 7000 kJ it prescribed either much less or no oily fish for the dietary models. Furthermore, there was little difference in the prescription of vegetables for each of the kilojoule targets, although the DMT prescribed up to 2 extra servings (up to 1 cup extra) than the other models.

## Discussion

### Principal Findings

The study described in this paper addressed the first-stage development of and comparative validation of an automated DMT. Applying lessons from previous linear modeling work, a dietary modeling tool using constraint optimization and nonlinear programming was developed online incorporating the expected energy requirement calculations in the same system rather than as separate data as per other less automated tools. Having the DMT online has the potential for improved user access, and creating an algorithm with constraints applied to it should also minimize the variability by comparison with manual modeling approaches. The identified differences between the approaches were not found to be clinically relevant. Although it was not the focus of this paper, it is likely that improved time efficiencies were created by automation of the DMT because of the incorporation of the EER calculation into the user interface. Furthermore, a reduced need for “guess work,” as is common to manual approaches, is also a significant advantage.

When comparing the dietary models created by the APD as the reference for comparison, and the models created by the less experienced dietitian student (manual models), the results across each of the kilojoule increments were most similar. It is likely that because of their training both users were aware of practical servings of particular food groups. Resultantly, excessive or limited servings of certain food groups were not identified in these 2 methods. Furthermore, being trained in the field of nutrition and dietetics and working closely with the Australian Dietary Guidelines, the diet models created by both users were most consistent with these recommendations. The partially automated models produced the next most similar outcomes to the reference models, although they still produced acceptable output for the majority of the kilojoule increments based on the Australian Dietary Guidelines. A benefit of using the Solver add-on was the restrictions that were able to be placed on the variability of energy coming from the macronutrients, ensuring the maximum reliability of individualized diets within a trial. Where possible, these were set within 3 percentage points of the macronutrient targets specified across each of the tools (50% carbohydrate, 20% protein, and 30% total fat). The DMT was also able to suggest appropriate dietary models for the majority of the kilojoule increments within these set limits. This was particularly evident for the higher kilojoule targets. The reference model and manual approach were not able to maintain such limited variability. The variability of the percentage of energy from macronutrients in the DMT fluctuated. For the majority of the models the variation between macronutrient contribution and the target was smaller than that seen in both the reference and manual models. The DMT output demonstrated the greater variation from the reference models; however, it must be noted that models could not be created



XSL•FO
RenderX

based on the exact kilojoule increment required for comparison as the other 3 models, as the DMT results were created based on a subject's *exact* EER rather than rounded to the nearest 500-kJ increment, demonstrating that it could be tailored to each individual participant of a clinical trial. The manual and partially automated methods were not tailored specifically to each individual because of the time-consuming nature of obtaining the modeling outcomes. Therefore, the outcomes gathered were closest to the required kilojoule increments. For a dietary intervention trial, diet models that have been specifically designed to meet the individual requirements of a participant are more desirable and, consequently, the method employed by the DMT would be preferable over the comparative methods in this study.

The largest variation across the dietary models was seen first, for the polyunsaturated fatty acids and second, for the monounsaturated fatty acids food groups. Interestingly, there is no exact recommendation in the Australian Dietary Guidelines in terms of servings related to these nutrients that could be referenced to create the models. It can be seen that these fats were largely used to make up the difference in energy and percentage of energy from fats required once all other food groups had been assigned, potentially explaining the difference. They also flag the need for a modification to the algorithms when nutrients beyond the macronutrient level are to be considered.

## Limitations

Challenges arose with the partially automated method and the DMT in terms of gaining appropriate outcomes. The Solver had to be tested with a range of constraints until desirable outcomes were being achieved across the kilojoule increments, again taking additional time until the desired result was achieved. When too many restrictive constraints were added, the tool was unable to compute a result. Therefore certain limitations (particularly leniency with the degree of variability from the set macronutrient targets) had to be made more liberal. The dietary modeling framework described in this paper had a number of developmental challenges to be overcome in order to produce comparative data. This included having the same results produced for each participant, irrespective of significant differences in age, height, and weight. Second, the outcomes produced were excessive in certain food groups (prescribing up to 7 servings of fruit daily) and limited in others (food groups contributing fatty acids, which were adjusted to be more liberal). Each challenge was overcome by modifying the constraints of the algorithm. Furthermore, serving sizes were altered to those traditionally recommended in dietetics, such as half a cup of cooked vegetables and 30 g of cheese for a serving, to make the tool more practical for use in food-based trials rather than based directly on the Excel calculator, which was developed based on exchange lists. A final alteration required the number servings of food groups to change depending on the EER of the participant rather than manipulating the frequency (in hours) that a food group should be consumed. The latter method assumed that individuals eat continuously over a 24-hour period, a case where professional judgment was needed to adjust the algorithm. Again, overcoming these issues when validating the DMT has ensured it is much more practical and simple to use and demonstrated that it can produce dietary models for macronutrient-based targets. It does, however, emphasize the importance of a nutrition expert, using the professional judgment [31], and working alongside computer programmers when developing tools for use in practice. Some limitations that could not be overcome within the current project include the inclusion of key nutrients beyond macronutrients. Inclusion of key nutrients in DMT, beyond the current macronutrients, will allow for a wider application to practice. The current form limits the use of the tool to studies related to overall energy constraints such as weight management or diabetes. Studies that aim to control dietary intake of vitamins and minerals could not use the tool in its present form. Furthermore, there is still a degree of professional judgment required with regard to the development of diet models. Complete individualization for food allergies or intolerance is not possible, although the researchers believe care professional judgment of appropriately qualified professionals such as dietitians should still be maintained in this instance.

## Comparison With Prior Work

It was also found that despite the type of modeling used in this study, careful consideration needed to be given to the rounding up or down of target servings. This was also identified in the AFGS [20] with rounding found to have minimal effect on the total energy being recommended. For the models of this study, models were largely rounded up to the nearest whole integer, although an underlying assumption was held that the final values would be used to create practical advice. Furthermore, the translation of food information from numbers determined in a model to practical suggestions provided by a dietitian in practice can vary significantly. This was also noted in the AFGS models as an additional consideration that needs to be addressed when liaising with an individual. Food allergies, food intolerances, and food preferences are only 3 of these considerations and are separate from those such as economic shifts, which may affect willingness to buy or purchase particular meat cuts, for example, or sustainable produce in relation to fish intake. Although DMT was developed to minimally address food preferences at a food group level, considerations in addition to this do require professional input and may not be appropriate to automate. Limitations were also apparent in comparing the data as certain food groups varied across the modeling methods. The DMT grouped fish with the other meats and protein-containing foods as is evident in the Australian Dietary Guidelines, instead of being a separate group as the other models. For the purpose of this study, to translate the information into practical suggestions for a participant and for the comparison to be made, these foods were separated by the dietitian into an appropriate amount of lean meats (based on what was prescribed by the tool) and the remaining portion considered as the fish.

## Conclusions

Dietary modeling is essential for the formulation of food-based prescriptions and useful to standardize background diets within randomized controlled trials. In its present form DMT provides this by using predefined macronutrient proportions for all participants of a dietary trial at the point of intervention. As demonstrated in this study, different dietary modeling tools with



XSL•FO
RenderX

the same dietary targets produced similar results. Manual methods for dietary modeling are less ambiguous in terms of the desired outcomes, as the model creator is aware of quantity prescriptions classified as being "appropriate." This method, however, requires trained dietary professionals to be able to produce desired results. When creating individualized prescriptions this can be a time-consuming process. Partially or fully automated methods such as the use of Solver and the developed DMT have the potential to be practically applicable for widespread use in dietary research. The DMT was found to be a valid automated tool producing similar results to tools with less automation. Once the underlying constraint systems have been formulated appropriately, use of such tools may not require trained professionals or those familiar with the Australian Dietary Guidelines for the development of all models, saving this expertise for practical translation of the models. This has significant resource implication for a research trial and even more so when considering the time saved when compared with the manual approaches used for dietary modeling. Future refinements are necessary to consider other nutrients such as key vitamins and minerals in the models to increase the flexibility of the tool and widen its application to practice. Further to this, inclusion of prompts within DMT related to food allergies or intolerances could also be included for further refinement of the model with a qualified practitioner.

Partially automated approaches such as that of Solver still require increased time to set the constraints; however, they will not produce results that are practically viable. Although by comparison with the manual method Solver does save some time, it continues to require the input of a professional to ensure the models produced are realistic. It is observed that although time was not monitored as part of this study, increased automation appears to relate to an increase in time saved and may in turn result in a reduction in resource allocations to the trial. With respect to the DMT the results of this study suggest interchangeability of the approaches, although implementation should reflect the requirements of the trial in which it is used and the available resources that can be used.

## Acknowledgments

The authors wish to thank Professor Jan Wright and Professor Peter Robinson for their intellectual contributions to the editing of the manuscript and the HealthTrack research team who provided input on the manual modeling process.

## Conflicts of Interest

None declared.

## Multimedia Appendix 1

Screen shots of the Dietary Modelling tool showing the study targets, patient details and food group output screen available from http://dietmodels.com/dmt.

[PDF File (Adobe PDF File), 1MB - jmir_v18i7e190_app1.pdf ]

## References

1. Dong Y, Hoover A, Scisco J. Muth E: A new method for measuring meal intake in humans via automated wrist motion tracking. Appl Psychophysiol Biofeedback 2012;37(3):205-215.
2. Probst Y, Zammit G. Predictors for Reporting of Dietary Assessment Methods in Food-Based Randomized Controlled Trials over a Ten Year Period. Crit Rev Food Sci Nutr 2016;56(12):2069-2090.
3. Cinthya W, Yasmine P, Neale E, Tapsell L. Impact of food supplementation on weight loss in randomised-controlled dietary intervention trials: a systematic review and meta-analysis. Br J Nutr 2016;115(8):1406-1414. [doi: 10.1017/S0007114516000337] [Medline: 26888153]
4. Wilson P, Dainty JR. Modelling in nutrition: an introduction. Proc Nutr Soc 1999;158(1):133-138.
5. Vyth E, Hendriksen M, Roodenburg A, Steenhuis I, van RJ, Verhagen H, et al. Consuming a diet complying with front-of-pack label criteria may reduce cholesterol levels: a modeling study. Eur J Clin Nutr 2012;66(4):510-516.
6. Food Standards Australia New Zealand. Food Standards. Canberra, Australia; 2009. Principles Practices of Dietary Exposure Assessment for Food Regulatory Purposes URL: http://www.foodstandards.gov.au/publications/Pages/Principles-and-Practices-of-Dietary.aspx [accessed 2016-07-05] [WebCite Cache ID 6ilZ9jjzO]
7. Dietitians Association of Australia. Eat for health. 2011. A modelling system to inform the revision of the Australian Guide to Healthy Eating URL: https://www.eatforhealth.gov.au/sites/default/files/files/public_consultation/n55a_dietary_guidelines_food_modelling_111216.pdf [accessed 2016-07-05] [WebCite Cache ID 6ilbd1zEG]
8. Food Standards Australa Zealand. Principles Practices of Dietary Exposure Assessment for Food Regulatory Purposes. Canberra, Australia: Food Standards Australia New Zealand; 2009.
9. Brehm B, Seeley R, Daniels S, D'Alessio DA. A randomized trial comparing a very low carbohydrate diet and a calorie-restricted low fat diet on body weight and cardiovascular risk factors in healthy women. J Clin Endocrinol Metab 2003;88(4):1617-1623.



XSL•FO
RenderX

10. Das S, Saltzman E, Gilhooly C, DeLany J, Golden J, Pittas A, et al. Low or moderate dietary energy restriction for long-term weight loss: what works best? Obesity (Silver Spring) 2009;17(11):2019-2024.
11. Kasim-Karakas S, Tsodikov A, Singh U, Jialal I. Responses of inflammatory markers to a low-fat, high-carbohydrate dietffects of energy intake. Am J Clin Nutr 2006;83(4):774-779.
12. Keogh J, Brinkworth G, Noakes M, Belobrajdic D, Buckley J, Clifton PM. Effects of weight loss from a very-low-carbohydrate diet on endothelial function and markers of cardiovascular disease risk in subjects with abdominal obesity. Am J Clin Nutr 2008;87(3):567-576.
13. Krebs J, Elley C, Parry-Strong A, Lunt H, Drury P, Bell D, et al. The Diabetes Excess Weight Loss (DEWL) Trial: a randomised controlled trial of high-protein versus high-carbohydrate diets over 2 years in type 2 diabetes. Diabetologia 2012;55(4):905-914.
14. Ross LA, Tapsell LC, Probst Y. Optimizing dietary fat in a weight-loss trial requires advice based on a structured whole-of-diet model. Nutrition Research 2011;31(9):683-690.
15. Tapsell L, Flood V, Probst Y, Charlton K, Williams P. Nutrition Tools: Dietary Assessment, Food Databases and Dietary Modelling. In: Food, Nutrition and Health. Melbourne, Australia: Oxford University Press; 2013:282-320.
16. Department of Health and Ageing. Eatforhealth. Canberra, ACT: National Health and Medical Research Council; 2011. A modelling system to inform the revision of the Australian Guide to Healthy Eating URL: https://www.eatforhealth.gov.au/sites/default/files/files/public_consultation/n55a_dietary_guidelines_food_modelling_111216.pdf [accessed 2016-07-05] [WebCite Cache ID 6iliR7jGE]
17. Probst Y, Morrison E, Dam HK. Dietmodels. 2015 Dec 15. The Dietary Modelling Tool URL: http://dietmodels.com/dmt/ [accessed 2015-12-18] [WebCite Cache ID 6dsRLrx05]
18. Gillen L, Tapsell LC. Development of food groupings to guide dietary advice for people with diabetes. Nutrition & Dietetics 2006;63(1):36-47.
19. Mackworth AK. Constraint satisfaction. In: Shapiro SC, editor. Encyclopedia of AI, Volume 1, 2nd Edition. New York: John Wiley & Sons; 1992.
20. National Health and Medical Research Council.: Department of Health and Ageing; 2014. Draft: A new food guideance system for Australia- Foundation total diets URL: http://consultations.nhmrc.gov.au/public_consultations/draft-new-food-guidance-system [accessed 2016-07-05] [WebCite Cache ID 6iljWZuIQ]
21. Briend A, Darmon N, Ferguson E, Erhardt JG. Linear programming: a mathematical tool for analyzing and optimizing children's diets during the complementary feeding period. J Pediatr Gastroenterol Nutr 2003 Jan;36(1):12-22. [Medline: 12499991]
22. Maillot M, Ferguson E, Drewnowski A, Darmon N. Nutrient Profiling Can Help Identify Foods of Good Nutritional Quality for Their Price: a Validation Study with Linear Programming. J Nutr 2008;138(6):1107-1113. [Medline: 18492842]
23. Tapsell L, Batterham M, Charlton K, Neale E, Probst Y, O'Shea J, et al. Foods, nutrients or whole diets: effects of targeting fish and LCn3PUFA consumption in a 12mo weight loss trial. BMC Public Health 2013;13:1231 [FREE Full text] [doi: 10.1186/1471-2458-13-1231] [Medline: 24369765]
24. Tapsell LC, Gillen LJ, Patch CS, Batterham M, Owen A, Baré M, et al. Including walnuts in a low-fat/modified-fat diet improves HDL cholesterol-to-total cholesterol ratios in patients with type 2 diabetes. Diabetes Care 2004 Dec;27(12):2777-2783. [Medline: 15562184]
25. Tapsell L, Batterham M, Huang XF, Tan S, Teuss G, Charlton K, et al. Short term effects of energy restriction and dietary fat sub-type on weight loss and disease risk factors. Nutr Metab Cardiovasc Dis 2010 Jun;20(5):317-325. [doi: 10.1016/j.numecd.2009.04.007] [Medline: 19570664]
26. Wheeler ML. Nutrient database for the 2003 exchange lists for meal planning. J Am Diet Assoc 2003;103(7):894.
27. Food Standards Australia New Zealand. Food Standards. NUTTAB Online 2006 URL: http://www.foodstandards.gov.au/consumerinformation/nuttab2006/ [accessed 2016-07-05] [WebCite Cache ID 6ilkYZ41G]
28. Mifflin M, St Jeor ST, Hill L, Scott B, Daugherty S, Koh YO. A new predictive equation for resting energy expenditure in healthy individuals. Am J Clin Nutr 1990 Feb;51(2):241-247. [Medline: 2305711]
29. National Health Medical Research Council. Dietary Energy. In: Nutrient Reference Values for Australia and New Zealand. Canberra, Australia: NH&MRC; 2003.
30. National Health and Medical Research Council. Australian Dietary Guidelines. Canberra, Australia: National Health and Medical Research Council; 2013. URL: https://www.nhmrc.gov.au/guidelines-publications/n55 [accessed 2016-07-07] [WebCite Cache ID 6ip8ZCdOe]
31. Probst Y, Tapsell L. What to ask in a self-administered dietary assessment website: the role of professional judgement. Journal of food composition and analysis 2007;20(8):696-703.


## Abbreviations

**AFGS:** Australian food guidance system
**APD:** Accredited Practising Dietitian
**CHO:** Carbohydrate



XSL•FO
RenderX

**DMT:** Dietary Modeling Tool
**EER:** estimated energy requirement
**NRV:** Nutrient Reference Value
**PA:** physical activity
**REE:** resting energy expenditure

*Edited by G Eysenbach; submitted 18.12.15; peer-reviewed by K Balter, B Davy, L Pao-Hwa; comments to author 24.01.16; revised version received 14.03.16; accepted 20.05.16; published 28.07.16.*

*Please cite as:*
*Probst Y, Morrison E, Sullivan E, Dam HK*
*First-Stage Development and Validation of a Web-Based Automated Dietary Modeling Tool: Using Constraint Optimization Techniques to Streamline Food Group and Macronutrient Focused Dietary Prescriptions for Clinical Trials*
*J Med Internet Res 2016;18(7):e190*
*URL: http://www.jmir.org/2016/7/e190/*
*doi:10.2196/jmir.5459*
*PMID:27471104*

©Yasmine Probst, Evan Morrison, Emma Sullivan, Hoa Khanh Dam. Originally published in the Journal of Medical Internet Research (http://www.jmir.org), 28.07.2016. This is an open-access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/2.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on http://www.jmir.org/, as well as this copyright and license information must be included.



XSL•FO
RenderX

Original Paper

# An Integrated Process and Outcome Evaluation of a Web-Based Communication Tool for Patients With Malignant Lymphoma: Randomized Controlled Trial

Inge Renske van Bruinessen[1], MSc; Evelyn M van Weel-Baumgarten[2], MD, PhD; Hans Gouw[3]; Josée M Zijlstra[4], MD, PhD; Sandra van Dulmen[1,2,5], PhD

[1]NIVEL Netherlands Institute for Health Services Research, Utrecht, Netherlands

[2]Radboud University Medical Center, Dept of Primary and Community Care, Nijmegen, Netherlands

[3]Hematon, Amersfoort, Netherlands

[4]VU University Medical Center, Dept of Haematology, Amsterdam, Netherlands

[5]University College of Southeast Norway, Drammen, Norway, Faculty of Health Sciences, Drammen, Norway

**Corresponding Author:**
Inge Renske van Bruinessen, MSc
NIVEL Netherlands Institute for Health Services Research
Otterstraat 118-124
Utrecht, 3500 BN
Netherlands
Phone: 31 302729680
Fax: 31 302729729
Email: i.vanbruinessen@nivel.nl

## Abstract

**Background:** The complex nature of the medical dialogue and the often emotional context in cancer care present challenges to health care professionals (HCPs) and patients. Patients are increasingly expected to be informed participants and to be able to make conscious decisions, which they often find very difficult. In an attempt to support patients with malignant lymphoma in clinical communication, we developed a stand-alone, Web-based intervention called "PatientTIME." The development of PatientTIME was based on a participatory intervention mapping framework. Its primary aim is to boost patients' self-efficacy in patient-professional communication (ie, their confidence when interacting with their HCP). Patients can use this intervention before their hospital visit to prepare for their clinical consultation. PatientTIME is fully automated and use is patient-initiated.

**Objective:** The aim of this study was to evaluate if and in what way patients benefit from PatientTIME and if it enhances their confidence in clinical communication.

**Methods:** The intervention was evaluated in a closed randomized controlled trial with continuous recruitment (using online and offline methods to reach potential participants) and data collection. In accordance with the Medical Research Council guidance, we started with a process evaluation. Subsequently, an outcome evaluation was performed focusing on the patients' perceived confidence in communication with their HCP, measured with the validated PEPPI questionnaire at baseline and at 3 months after participation. Process and outcome data were obtained through Web-based questionnaires, log files (automatically generated files mapping the interactions between program and users), and a logbook (comprising a record of actions and interactions kept by the researchers). Participants were not blinded. A total of 146 patients registered online, of whom 97 gave their informed consent and were assigned at random to the control group (N=34) or 1 of the 2 intervention groups (N=63). Ultimately 87/97 (90%) of these patients actually participated in the study, producing 87 datasets for analysis.

**Results:** More than half of the intervention group patients reported that the intervention helped them prepare for a clinical consultation; it created awareness about the importance of communication and reinforced their existing communication skills. In the postvisit test, the control group showed a small, nonsignificant improvement in perceived communication efficacy. The intervention group showed a significant improvement in perceived efficacy. However, the interaction effect was not significant, indicating that the improvement solely as a result of the intervention may not be significant.



XSL•FO
RenderX

**Conclusions:** A considerable number of patients reported that PatientTIME did provide support. We found a trend indicating that in the long run, patients with access to PatientTIME scored better on the perceived efficacy scale than patients without access. However, at this stage we cannot conclude that PatientTIME improves patients' confidence when interacting with HCPs.

**ClinicalTrial:** Netherlands National Trial Register (NTR): 3779; http://www.trialregister.nl/trialreg/admin/rctview.asp?TC=3779 (archived by WebCite at http://www.webcitation.org/6iztxJ5Nt)

***(J Med Internet Res 2016;18(7):e206)*** doi:10.2196/jmir.5877

**KEYWORDS**

RCT; communication aid; Web-based intervention; self-help application; hematologic malignancies; lymphoma cancer; patient participation

## Introduction

The interaction between the health care professional (HCP) and the patient is the fundamental vehicle for exchanging information. For the HCP, effective communication is necessary to manage and resolve biomedical and psychosocial problems, which are key issues in cancer care. For the patients it is important to "know and understand," and the communication serves a purpose in their need to "feel known and understood" [1].

It is important to have effective communication in order to deliver good care. Indirectly, effective communication has been linked to a range of improved patient outcomes such as satisfaction, treatment compliance, perceived quality of life, and physical health [2-6]. However, the complex nature of the medical dialogue and the often emotional context in cancer care are a challenge for HCPs and patients, and the quality of communication often remains suboptimal [7]. Although the HCPs are responsible for the communication process, the increased focus on patient empowerment and shared decision making has broadened the role of patients [8-10]. Patients are increasingly expected to be informed participants and to be able to make conscious decisions [11].

Research shows that such patient participation pays off: if patients participate actively, physicians provide significantly more information overall and respond better to questions [12]. Patients who reach their preferred level of participation experience less anxiety and are more satisfied with the clinical consultation [13]. However, most cancer patients do not achieve their desired level of participation [13,14]. Patients' communicative contribution appears to be limited [15,16] and patients report unmet communication needs [7,17]. Research has highlighted the importance of not only training the HCPs in communication skills, but also providing cancer patients with support in communication [18-20].

So far, cancer communication studies in clinical settings focus mostly on specific types of cancer, especially breast, prostate, and colorectal cancer [21]. Disease-specific communication instruments are lacking for patients with malignant lymphoma. Via the Dutch patient association Hematon (for leukemia, malignant lymphoma, and stem cell transplantation), these patients have indicated that they often lack the skills needed to be more in control, participate more, and play a more active role during clinical consultations. Research confirms their need for support [22,23]. In an attempt to support patients with malignant lymphoma in communicating with their health professionals, we collaborated with these patients to develop the Web-based intervention "PatientTIME" [24]. Patients can use this stand-alone intervention before their hospital visit to prepare for clinical consultations (see Intervention). The primary aim of the intervention is to positively influence patients' self-efficacy in patient-professional communication [25,26], that is, their confidence that they can interact with their HCP. Self-efficacy is an important predictor of actual communication behavior [21]. The effectiveness of PatientTIME was tested in a randomized controlled trial (RCT) with self-efficacy as the primary outcome measure.

Randomized controlled trials are considered to be the most rigorous way of evaluating effectiveness in the medical context. Traditionally, the main focus is on reporting prespecified outcomes. This evaluation method is predominantly applied in interventions with one active variable, for example, the effect of a drug on survival [27]. In interventions like PatientTIME, different active ingredients (Table 1) are combined and evaluated simultaneously. Oakley et al [28] argue that when evaluating such a "complex" intervention, incorporating a process evaluation would support and improve the interpretation of outcomes. Process evaluations look into the nature of the intervention, how it is delivered, and what actually happens during the intervention [29,30]. It can improve the validity and interpretation of outcomes, help refine the intervention, and provide necessary information for replication [27,30]. Despite the rise of complex interventions, few studies combine process *and* outcome evaluations.

In this study, knowledge about the process characteristics is expected to help in improving the PatientTIME intervention: it may show how to reach different patient groups and it can support the right interpretation of outcomes. Moreover, the process evaluation provides the context in which the data for the outcome evaluation are gathered. The main question to be answered by the outcome evaluation is "Does the intervention increase participants' confidence in interacting with their HCP?"

The ultimate goal is to implement PatientTIME as a publicly available, stand-alone intervention, that is, without the research context and without the involvement of professionals. In addition to giving insight into the effectiveness, the results of the study can help us optimize PatientTIME as a stand-alone intervention.



XSL•FO
RenderX

## Methods

### Procedure and Ethical Approval

In accordance with the Medical Research Council (MRC) guidance, we started with a process evaluation focusing on the reach of the intervention and the extent to which it was used as intended [31]. Subsequently, the outcome evaluation was performed, focusing on the patients' perceived confidence when interacting with their HCP. The research ethics committee of the Radboud University Nijmegen Medical Centre evaluated the RCT protocol and concluded that the study did not fall within the remit of the Dutch Medical Research Involving Human Subjects Act (WMO). The study is registered in the Netherlands National Trial Register (trial registration number NTR3779). Written informed consent forms were used.

### Intervention

The Web-based PatientTIME intervention aims to support patients in gaining more control over the communication with their HCP. The intervention development was guided by the intervention mapping framework applied in close collaboration with patients [24] and makes use of different theory-based methods: modeling, tailoring information, previsit goal setting, and listening to visit recordings. The central source of information in the intervention consists of 58 short video fragments (47-180 seconds) showing simulated patients demonstrating different communication skills during medical encounters (eg, stating the need for support, dealing with emotions, or asking questions; Figure 1) [32]. The fragments are based on communication barriers identified by the targeted population in a previous study [22]. A question prompt sheet (QPS) and an option to replay an audio recording of the user's hospital visit were also included in the intervention. The functionality and intended use of these individual components are described in Table 1. The collaborative partners (2 hospitals, the patient association, the funding organizations, and a research institution) were listed on an information page.

**Table 1.** The intended use of the individual intervention components.

| Intervention component | Intended use |
|---|---|
| Video library | Before a clinical consultation, a subset is selected from the 58 video fragments available for use in the video library. The selection is tailored to the user's preferences and needs at that time and stored in the user's personal video library. When the intervention is used again, new video fragments are added to the library along with the previously viewed videos (which are still available for viewing). Per consultation, video clips regarding a maximum of 3 communication themes are provided (6 clips in total). When using the intervention for the first time, a maximum of 4 introductory clips are added to the theme clips. |
| Question prompt | A prompt was integrated to encourage patients to set goals and prepare questions before the consultation. Patients can also formulate questions or remarks while watching the video fragments. The question prompt sheet can be printed or emailed to the patient's address. |
| Listening back to the consultation | A consultation audio recording can be uploaded, attached to the consultation date in the agenda, and replayed at any time. |

### Recruitment

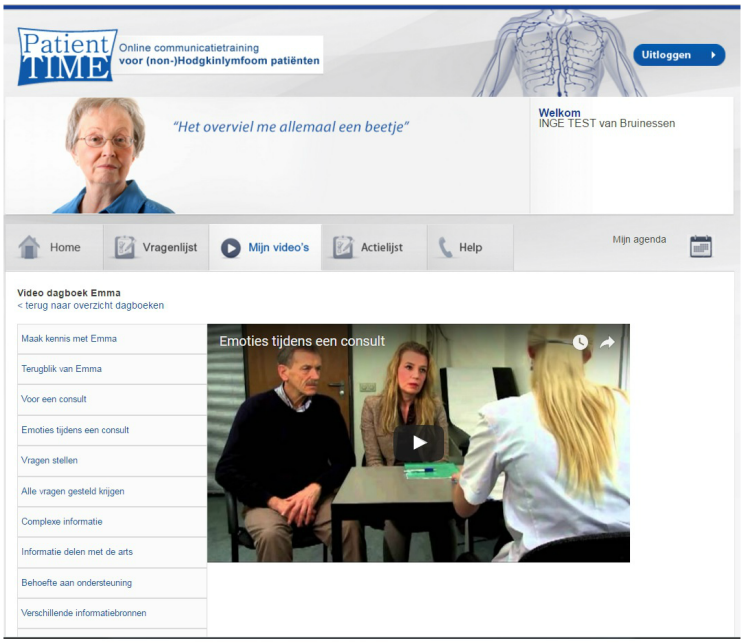
PatientTIME is a stand-alone intervention; that is, patients can register and use the intervention without referral or the involvement of a professional. Several online and offline actions were organized to create awareness of the availability of PatientTIME: (1) spreading leaflets at hospital waiting rooms, (2) giving short presentations during patient information meetings arranged by Hematon and at hospitals, (3) short news flashes in Hematon's digital newsletter and magazine, (4) distributing our own project newsletters, and (5) using social media (Twitter, forums). Additional actions were initiated during the inclusion period: (6) we sent information packages and leaflets to hematologists asking them to distribute the information among their patients, and (7) we sent short news flashes about the study to local newspapers. Patients could express their interest in participating via registration on the PatientTIME website [21] between March 2013 and May 2015. After registration, the study's inclusion criteria were verified. Patients were eligible for participation if they were aged 18 years or older, had been diagnosed with malignant lymphoma, had at least one (follow-up) consultation with their HCP per year, had a good understanding of the Dutch language, were receiving treatment or follow-up care in a Dutch hospital, and had access to a computer with an Internet connection.

If the inclusion criteria were met, patients received a consent form with additional information by post. As soon as they had returned the informed consent form, their account was activated and an email was automatically sent to the patient with log-in details. If no consent form was returned, the patient was reminded by email after 2 weeks and by phone 1 week after that.



XSL•FO
RenderX

**Figure 1.** Screenshot of the PatientTIME intervention.

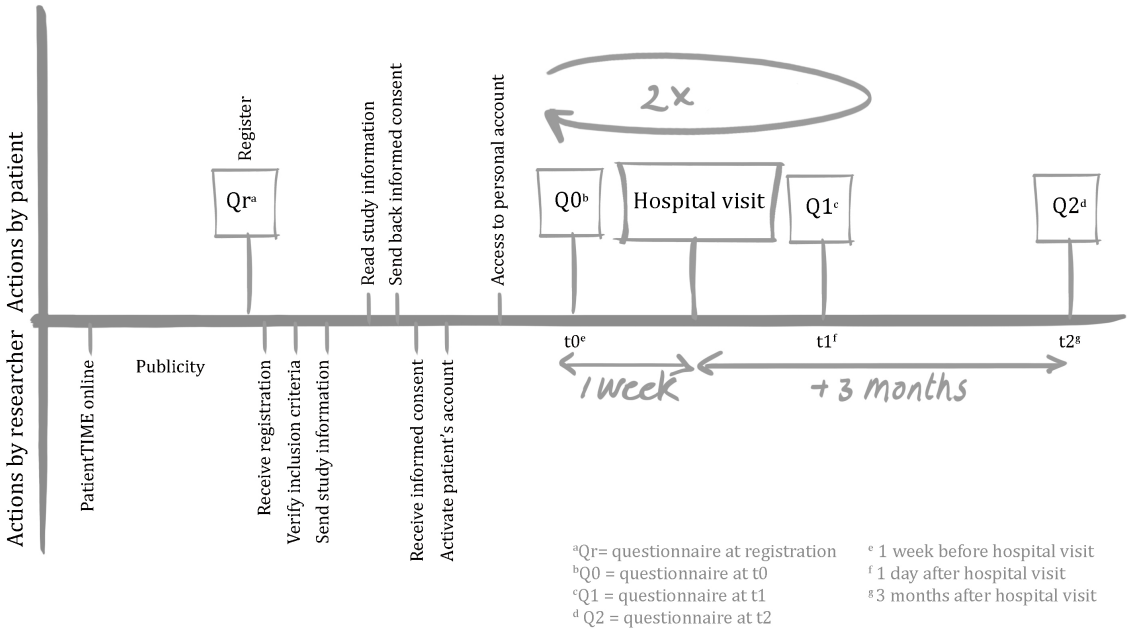



## Design

The intervention was implemented in the form of a 3-armed RCT with continuous recruitment and data collection. The computer assigned participants randomly to 1 of 3 groups: (1) a control group, (2) an intervention group, and (3) an intervention audio group. Software that enabled blocked randomization (block size 3) was used to balance the groups.

The 2 intervention groups differed in one component: having or not having the opportunity to listen afterward to the audio recording of their clinical consultation. For the purpose of this outcome evaluation, the 2 intervention groups were evaluated as 1 group and compared with the control group. We made this decision for practical reasons. Patients who had the opportunity to listen afterward to their consultation generally did this after completing the postvisit questionnaire (Figure 2). Therefore, the experience of replaying and listening to the consultation could not have influenced their answers. Moreover, some patients in the intervention group where we did not actively encourage them to record their consultations (group 2) still made recordings on their own initiative.

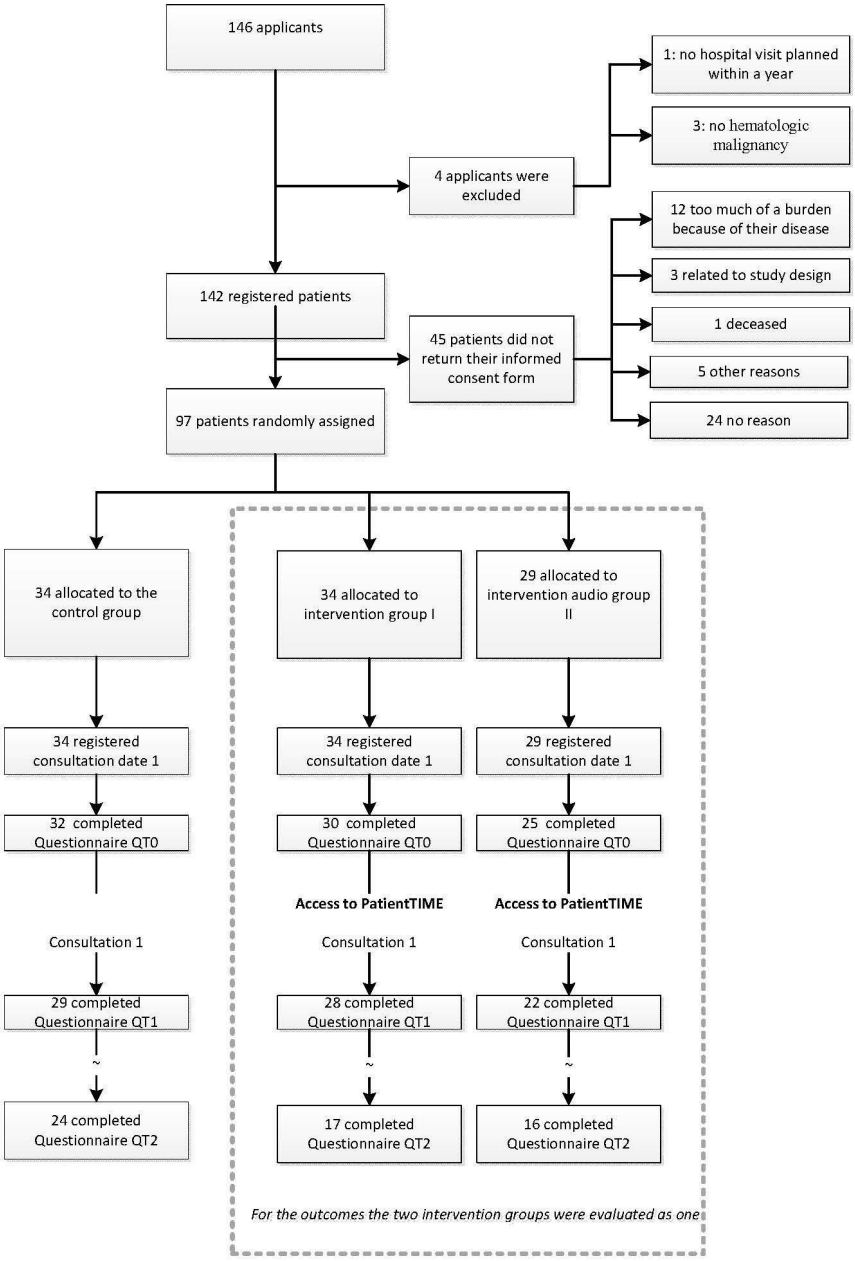
**Figure 2.** Timeline randomized controlled trial steps.






XSL•FO
RenderX

**Figure 3.** Flowchart.




All participants were given access to a personal secure Web account with questionnaires and an integrated digital agenda. The agenda had an automatic reminder system. This system led them through the study with personalized messages and questionnaires before and after clinical consultations (Figure 2). The date of the participant's forthcoming hospital visit determined the start of his or her participation. One week before this visit, participants received an automatic email reminder with the request to complete the previsit questionnaire on their Web account. After having completed the previsit questionnaire, the control group received no further information. The 2 intervention groups were given access to the intervention content (Figure 3). The selection of video fragments taken from the video library was based on the participant's self-reported previsit communication preferences and needs. The intervention audio group participants were also instructed—if permitted by their HCP—to record their consultations, upload the audio recordings, and listen afterward to them. The recordings were also available to the researchers for analyses. We provided these participants with a small audio-recording device and extra information about consultation audio recordings.

The day after the consultation, patients received an automatic reminder with a link to the postvisit questionnaire. If a second and third consultation was planned within the participation year, the cycle was repeated with less comprehensive questionnaires. If necessary, a maximum of 2 reminders were sent by the researcher reminding the participant to complete the questionnaires or register a consultation date. The control group had access to the educational content of the intervention after participation.

There was no downtime during the evaluation period. PatientTIME's content and functionality were frozen for the duration of the trial. Necessary bug fixes were made relating to questionnaires that were not always directly available after the clinical consultation.



XSL•FO
RenderX

Participants were not blinded as they could see whether they had access to the content of PatientTIME or not. As self-confidence in communication was the primary outcome measure, the power calculation was based on the primary outcome measure of the RCT, the PEPPI (Perceived Efficacy in Patient-Physician Interactions) questionnaire (see Data Collection and Measures) [26]. We expected the intervention to result in a decrease of 5 points on the sum score (reflecting higher confidence), which corresponds to an effect size of 0.70. To detect this difference with a power of .80 and alpha of .05, we needed 50 patients in the intervention group and 50 patients in the control group.

## Data Collection and Measures

The process and outcome data were measured through (1) Web-based questionnaires (Qr, Q0, Q1, and Q2; see Figure 2), (2) log files, that is, automatically generated files mapping the interactions between the program and its users, and (3) a logbook, comprising a record of actions and interactions that was kept by the researchers.

According to the MRC guidance, the selection of dimensions that are evaluated should be adjusted to the intervention under study. A total of 7 process dimensions were selected that may influence outcome measures in this study (Table 2).

**Table 2.** Dimensions of the PatientTIME process evaluation.

| Dimensions | Purpose | Research questions (data source) |
|---|---|---|
| **Delivery** | | |
| Uptake | Insight into the procedures used to reach users, and information on who was reached | Were the publicity actions conducted as planned? (Logbook)<br>Who showed interest in participating? (Questionnaire Qr)<br>What were the reasons for not participating? (Logbook) |
| "Dose" delivered | Insight into the exposure to the intervention | To whom was the intervention delivered, and how does this compare to the population of interested patients? (Questionnaire Qr) |
| Participation rate | Insight into the characteristics of participants[a] and their reasons for participating | Who participated in the evaluation, compared with the total group who were invited to participate in the intervention? (Questionnaire Qr, Q0) |
| Fidelity | Insight into the extent to which the intervention was provided as planned | Which adaptions made to the intervention for the benefit of the effect evaluation may have influenced use and outcomes? (Logbook)<br>What was the time between registration and the consultation date? (Questionnaire Qr)<br>To what extent were reminders necessary for the benefit of the effect evaluation? (Logbook) |
| **User-program interaction** | | |
| Dose received (attrition) | Insight into the actual exposure to the intervention and the use of its different components | How many video fragments were watched per patient? (Log files)<br>How many patients used the question prompt sheet? (Log files)<br>How many patients audio-recorded their consultation? (Log files) |
| Usability | Insight into the usability of the intervention | How was the satisfaction with the intervention rated? (Questionnaire Q1)[b] |
| **Contextual factors** | | |
| Patient's context | Insight into the contextual factors at the micro level | What influence does the patients' context have and how could the context have influenced the outcome evaluation? (Logbook) |

[a] A patient was defined as a "participant" if he or she registered the first consultation date and completed the previsit questionnaire Q0.

[b] The usability was evaluated with the System Usability Scale, a 10-item questionnaire that gives an overview of satisfaction with the program, resulting in a sum score (range 0-100) [33]. Usability tests had already been conducted with patients on a small scale during the development phase. However, as usability is strongly associated with use, attrition, and dropout, we decided to measure it on a larger scale as well.

In addition to the dimensions taken from the MRC guidance, the intervention group was asked if and in what way the intervention had helped them. All "evaluative" responses (answers to the question) were coded as "positive" or "negative." To prevent subjective interpretation of the data, 2 researchers coded all fragments. The interrater reliability was good (81% agreement). Disagreements were discussed to come to an agreement.

For the primary outcome measure, participants were asked to complete the validated, 10-item PEPPI questionnaire twice: at t0 and t2. The PEPPI questions all begin with "How confident are you in your ability to..." and assess a patient's confidence in their ability to communicate with their HCP on a 5-point Likert scale (1 = very confident to 5 = not confident at all) [26,34]. Item scores were summed (giving a range of 10-50), whereby lower scores reflect higher self-efficacy. The PEPPI questions used at t1 were reformulated so that they could also be used as a postvisit scale, measuring efficacy experienced directly after the visit.

## Statistical Analysis

Descriptive statistics were used to characterize the sample and to analyze the process measures.

Differences between groups were analyzed using *t* tests and chi-square tests, where appropriate. To analyze short-term



XSL•FO
RenderX

effects, linear regressions were conducted to test for differences in experienced efficacy measured at t1 between the control and intervention groups, adjusting for the previsit PEPPI score. To analyze long-term effects, multilevel regressions were conducted to test for differences on the PEPPI scale between the control and intervention groups at t2 and t0. The analysis of the open questions was performed using MAXQDA 11 software [35]. Other process and outcome data were analyzed with Stata 13.

## Results

### Participant Flow

A total of 146 potential participants registered (*uptake*) of which 97/142 (68%) patients met the inclusion criteria and signed the informed consent form (*dose delivered*; Figure 3). This group included 5 patients who were not diagnosed with malignant lymphoma but with a hematologic malignancy with a comparable impact (eg, chronic lymphocytic leukemia or multiple myeloma).

Of the 49 potential participants who were not included, 4 were excluded by the researchers. The remaining 45 did meet the inclusion criteria but did not return the informed consent form. Compared with these 45 patients, the 97 participants had a significantly higher level of education, were more likely to be a member of a patient organization, had a longer elapsed time between diagnosis and registration, and were more likely to have had or started treatment (Table 3).

All 97 patients in the RCT registered their first consultation date, which was a precondition for being sent the previsit questionnaire (Q0). Subsequently, the *participation rate* was 90% as 87/97 patients completed Q0 and were marked as participants. Compared with the nonparticipants (N=10), the 87 participants had a higher level of education and used the Internet on a more regular basis. After having completed Q1, a total of 76 participants registered a second consultation date and 52 participants a third consultation date as well. There was no significant difference between the control and intervention groups in the proportion registering multiple consultations. Because not every participant had a second or third consultation planned within the participation time span (1 year after registration), the analysis of the outcomes is based on the first consultation only.

### Process Evaluation

The identified process results that facilitated correct interpretation of the outcomes were part of the delivery domain (eg, fidelity, reach; see Table 2) and user-program interaction domain (eg, exposure, use of different components; see Table 2), in addition to the benefits perceived by the patients.



XSL•FO
RenderX

**Table 3.** Background characteristics and outcomes.

| Variable | Value | Registered N=142 | Randomized N=97 | Participant N=87 | Control N=32 | Intervention N=55 |
|---|---|---|---|---|---|---|
| Age (at registration) | Mean in years (SD) | 57 (14) | 55 (13) | 56 (13) | 56 (14) | 55 (13) |
| Sex, n(%) | Male | 63 (44) | 40 (41) | 34 (39) | 14 (44) | 20 (36) |
| Education[a], n(%) | Low | 16 (11) | 6 (6) | 2 (2) | 2 (6) | 0 |
| | Medium | 39 (27) | 23 (24) | 19 (22) | 7 (22) | 12 (22) |
| | High | 89 (61) | 68 (70) | 66 (76) | 23 (72) | 43 (78) |
| Member of a patient organization[b], n(%) | Yes | 71 (50) | 60 (62) | 56 (64) | 16 (50) | 40 (73) |
| | No | 70 (49) | 37 (38) | 31 (36) | 16 (50) | 15 (27) |
| Internet use[c], n(%) | Daily | 128 (90) | 90 (93) | 90 (93) | 29 (91) | 52 (95) |
| | Weekly | 10 (7) | 6 (6) | 6 (7) | 3 (9) | 3 (5) |
| | Monthly | 2 (1) | 1 (1) | 0 | 0 | 0 |
| **Disease-related information** | | | | | | |
| Diagnosis, n(%) | Hodgkin lymphoma | 31 (22) | 22 (23) | 20 (23) | 5 (16) | 15 (27) |
| | Non-Hodgkin lymphoma | 105 (74) | 70 (72) | 63 (72) | 26 (82) | 37 (67) |
| | Other | 6 (4) | 5 (5) | 4 (5) | 1 (3) | 3 (5) |
| Type[b], n(%) | Indolent | 57 (40) | 38 (39) | 35 (40) | 13 (41) | 22 (40) |
| | Aggressive | 41 (29) | 26 (27) | 23 (26) | 8 (25) | 15 (27) |
| | Combination | 13 (9) | 10 (10) | 8 (9) | 4 (13) | 4 (7) |
| | Not known (yet) | 30 (21) | 23 (24) | 21 (24) | 7 (22) | 14 (25) |
| Time elapsed since diagnosis | Mean in years (SD) | 5.4 (7.7)[c] | 6.5 (8.3)[b] | 6.2 (5.5)[b] | 3.2 (3.7) | 8.0 (10.0) |
| | < 1 year, n(%) | 42 (30) | 24 (25) | | | |
| | > 1 year, n(%) | 100 70) | 73 (75) | | | |
| Treated, n(%) | Yes | 121 (86) | 89 (92) | 79 (91) | 4 (13) | 4 (7) |
| | No | 20 (14) | 8 (8) | 8 (9) | 28 (88) | 51 (93) |
| Current status[b], n(%) | Awaiting test results or treatment (plan) | 19 (13) | 9 (9) | 9 (10) | 4 (13) | 5 (9) |
| | Currently getting treatment | 33 (32) | 22 (23) | 19 (22) | 8 (25) | 11 (20) |
| | Wait-and-see | 36 (26) | 26 (27) | 25 (29) | 9 (28) | 16 (29) |
| | Remission | 53 (38) | 40 (41) | 34 (39) | 11 (34) | 23 (42) |
| **Confidence in communication (PEPPI[d])** | | | | | | |
| Perceived, previsit (t0) | Range 10-50 (Cronbach alpha .94) | N/A[e] | N/A | 20.5 (6.3) | 20.3 (6.0) | 20.7 (6.4) |
| Experienced, postvisit[f] (t1) | Range 10-50 (Cronbach alpha .93) | N/A | N/A | 16.5 (7.8) | 15.9 (6.7) | 16.9 (8.3) |
| Perceived after > 3 months[g] (t2) | Range 10-50 (Cronbach alpha .93) | N/A | N/A | 19.3 (5.4) | 19.2 (6.4) | 19.3 (4.7) |

[a] Classification according to the guidelines of Statistics Netherlands.

[b] One missing.

[c] Two missing.

[d] PEPPI: Perceived Efficacy in Patient-Physician Interactions.



XSL•FO
RenderX

[e] N/A: not applicable.

[f] Ntotal=78, Ncontrol=28, Nintervention=50.

[g] Ntotal=57, Ncontrol=24, Nintervention=33.

Regarding the delivery domain, we concluded that the publicity actions were conducted as planned. On the whole, the intervention was indeed used as a stand-alone program; most participants did not need any help or reminders from the researchers. The questions they asked (mainly by email) concerned problems with logging in or study questionnaires that were not yet accessible. For the first consultation, 18 of the 97 randomized patients received at least one reminder to complete their previsit questionnaire.

We noted one issue regarding the "fidelity" that may have influenced outcomes. At registration, patients had to provide the date of their upcoming consultation. Of the potential participants who did this, 44/138 (32%) had their consultation planned within a fortnight. These patients were less likely to participate in the study (59%) compared with the patients who had their consultation later (76%).

Regarding the user-program interaction domain, we report the actual use of the different intervention components that may have influenced the outcomes. Of the participants assigned to the intervention group, 55/63 (87%) completed Q0 and had access to PatientTIME. This group viewed on average 6 (SD 3.5) of the 10 selected video fragments before their first consultation. Of all the fragments they watched, 76% were viewed from beginning to end. The QPS was used to write down questions, physical complaints, and concerns by 20/55 (36%) of the participants who had access to it. Of the intervention audio group 9/29 (31%) participants reported after consultation that they had recorded their visit, 13/29 (45%) reported that they had not, and 7/29 (24%) did not complete this question. Of the participants who did not record their consultation, 8/13 said that they decided that they did not want to make the recording, 2/13 said the HCP did not give permission, 2/13 said they did not want to ask the HCP and 1/13 forgot the recording device.

Regarding the "usability," 50 intervention group participants completed the System Usability Scale questionnaire at Q1, revealing an average satisfaction score of 73 points (SD 16) on the 0-100 scale, which is considered "good" according to the study by Bangor et al [36].

As for perceived benefits, in the intervention group, 46/50 (92%) participants completed the open question "Did you benefit from the website?" at t1. The positively coded answers (59%) explained that (1) the website was instructive, insightful, or helpful for the preparation of a clinical consultation; (2) the website created awareness about the importance and role of communication; and (3) participants recognized the scenarios, which gave them the feeling that they were doing all right.

> *I've been thinking somewhat more about the questions I was going to ask the medical specialist.*
>
> *It makes you more aware of yourself, but also of the doctor's role.*

The negatively coded answers (41%) referred to (1) bugs in the questionnaire or usability problems in the website and (2) the fact that the information did not add anything to what was already known or experienced.

> *Not that much. I have the idea that I am already quite outspoken during the consultations with my oncologists.*
>
> *Not much because I have a good relationship with my physician. I can discuss everything.*

### Outcome Evaluation

The results on the PEPPI scale at t0 and t2 (Table 3) were skewed to the lower scores, reflecting a group of patients with a high level of self-confidence about interacting with their HCP. The same skewed trend was visible at t1, reflecting a group of participants who were well able to reach their goals.

There were no differences on the PEPPI scale between the control and intervention groups at baseline (t0, $P$=.78). Directly after the consultation (t1), there was no difference between the control and intervention groups in the way they experienced their efficacy during the consultation (controlling for the previsit PEPPI score). The results remained the same when controlling for sex, the level of education, and time since diagnosis.

Comparing the PEPPI score at t2 with the PEPPI score at t0, the control group showed a small improvement in the level of perceived efficacy. This improvement (−0.38 points) was not significant ($P$=.69). The intervention group also showed an improved level of perceived efficacy. This improvement (−1.97 points) was significant ($P$=.02). The interaction effect (−1.59), however, was not significant ($P$=.20), indicating that the improvement solely as a result of the intervention may not be significant. In addition to this primary analysis, we did not find an association between the PEPPI outcomes and the number of video fragments watched.

## Discussion

### Overview

In this paper, we described an integrated process and outcome evaluation of the Web-based intervention PatientTIME. PatientTIME was developed with the aim of helping patients with their communication during clinical consultations. With the process evaluation, we aimed to sketch the context in which the outcome evaluation was performed. The process evaluation was also used to obtain input on how to improve the intervention and its reach before making the intervention publicly available. In the outcome evaluation, we focused on the patients' perceived confidence in interacting with their HCP.

### The Process Evaluation

As a result of the process evaluation, we identified a substantial number of delivery and interaction aspects that will help to improve the functionality of PatientTIME and its reach. Looking at the delivery, we want to highlight the recruitment process.



XSL•FO
RenderX

The chosen method may resemble how patients are informed in practice about publicly available Web-based interventions, but despite the various publicity actions during the recruitment period, this approach did not appear to be enough to recruit a large, diverse group of patients for the study (see also Trial Participants section).

The patient-program interaction evaluation showed how the different PatientTIME elements (video fragments, QPS, and consultation audio recordings) were used. The core content of the intervention, the video fragments, was well used. The percentage of fragments viewed entirely (76%) may even be an underestimate (fragments were only logged as "ended" when the screen switched back to start).

The QPS was used less frequently, and slightly less compared with other studies that evaluated the use of open QPS tools in oncology care [37,38]. We do not know if the limited use was due to the design of the QPS, the fact that patients found it difficult to articulate questions, or the fact that patients preferred to make their notes on paper. Albada et al found that patients need to be motivated to use a QPS. Information about why and how to use the QPS could have been made more explicit in the PatientTIME intervention [37]. In the intervention audio group, more than half of the participants did not record their consultation on audio. There appeared to be a variety of obstacles. Other studies show that cancer patients highly value consultation audio recordings and that the majority benefit from listening to the consultation afterward [39,40]. However, in these studies the HCP facilitated the recording, whereas in our study patients were encouraged to take the initiative. We may have failed to provide sufficient information to remove existing obstacles, or the fact that it was part of a trial may have caused patients to be reluctant. To conclude, the QPS and audio recording options should be improved in the PatientTIME intervention.

More than half of the intervention group patients explained that the intervention helped them prepare for a clinical consultation, created awareness, and/or reinforced their existing communication skills. Patients who reported no benefit explained that they already had a good understanding with their HCP, had a lot of experience, or considered themselves as sufficiently participative and empowered already.

## The Outcome Evaluation

Looking at the data on confidence in interaction, we found a trend indicating that in the long run, patients with access to PatientTIME scored better on the perceived efficacy scale than patients without access. Differences were, however, small and we did not reach our aim of an improvement of 5 points. Therefore, we cannot conclude at this stage that the intervention positively influenced participants' confidence in their ability to talk with their health professionals.

## The Results in Perspective

The strengths and limitations related to the study and intervention put the results in perspective and illustrate our lessons learned. We identified four important aspects.

### *Trial Participants*

The participants were mainly more educated, experienced patients who were already quite confident in their ability to talk with HCPs. It is a well-known phenomenon in eHealth research that patients with a high level of education are overrepresented in these studies [41,42]. However, a recent study in the Netherlands revealed that the majority of health care users are using the Internet to get information about care and health [43]. This makes it reasonable to assume that a broader group will find PatientTIME when it becomes publicly available. For the outcome evaluation, it is therefore a limitation that less experienced patients (patients who had only recently been diagnosed) and patients with a low initial confidence in their ability to interact were not well represented.

The possible explanations for this are (1) we did not reach them, (2) the idea of a communication support tool did not appeal to them, or (3) they did not want to be involved in a research project. Insight into the reach of the intervention is limited because of the lack of information about patients who read about the intervention but decided not to register for participation. The second explanation seems unlikely given the findings of a recent study among patients with a chronic illness (including cancer patients) that revealed that a considerable number (39%) are interested in communication support [44]. Participating in a study like this demands much more from patients than only using the PatientTIME intervention. We received many questions about the consultation audio recording in particular and patients had concerns about this component. All these extra elements may have made patients reluctant to take part.

### *Outcome Measure*

The participants in this study appeared to be highly confident in their ability to talk with the HCP at baseline, which left little room for improvement (ceiling effect). On the one hand, this suggests that we reached a group who largely did not need support in communication. On the other hand, these patients did enroll for the study and the core information in the intervention tool was well used. This indicates that these patients were interested in improving their communication skills. Perhaps they saw a mismatch between their preferred role and their behavior in the consultation room. This casts doubts on our decision to measure effectiveness with the PEPPI questionnaire only. Bensing et al reported that patients' behaviors in the consulting room are not necessarily a reflection of their self-reported preferred behavior [45]. Because we did not observe the consultations, we cannot describe participants' actual behavior or how PatientTIME may have influenced this. Although the PEPPI questionnaire is a good measure for tracking confidence in medical communication, observations of actual communicative behavior can be a valuable addition. Moreover, the lack of power limits the usefulness of the PEPPI results, as the control group did not reach the required 50 participants. It would also have been interesting to have the patients complete the PEPPI questionnaire again immediately after viewing the PatientTIME content, to measure if their level of confidence had changed as an immediate result.



XSL•FO
RenderX

### *The One-Sided Intervention Approach*

PatientTIME was solely focused on patients, without the interference of or intervening in the communication style of the HCPs. Butow et al found that their patient-targeted communication intervention was only effective when the patient consulted a clinician who endorsed the intervention [46]. The one-sided approach used in PatientTIME may have decreased the potential reach and effect.

### *Implicit Trial Effects*

Two trial effects were identified that may have influenced outcomes. First, regardless of whether participants were allocated to the intervention group or the control group, it is likely that the mere fact of participating in a communication study highlighted the importance of the participant's role in medical communication. Increased awareness of being observed in a trial setting may have affected outcomes (pre- and postvisit), which limits generalizability (Hawthorne effects).

Second, the relatively lengthy time between registration and getting access to the intervention tool (because study information needed to be provided and informed consent obtained) delayed the inclusion process. This may have excluded patients who were looking for communication support shortly before their planned consultation. Ideally, access should be given immediately.

## Conclusions

A considerable number of patients indicated that the intervention was helpful in preparing for a clinical consultation, created awareness about the importance of communication, and reinforced their existing communication skills. These are valuable findings as such. However, at this stage we cannot conclude that PatientTIME improves patients' confidence when interacting with health professionals. By integrating a process evaluation and an outcome evaluation in this way, we were able to demonstrate which elements of PatientTIME were used and appreciated, even though they did not lead to measurable changes in communication self-efficacy.

## Future Research

We identified three themes for future research. First, when evaluating the effectiveness of stand-alone Web-based interventions like PatientTIME, it may be worthwhile to consider research methods that allow the intervention to be dynamic, rather than freezing it in a certain state. In practice, Web-based interventions also need to be optimized continuously. With a more dynamic approach, intermediate test results can be used to improve the intervention and new versions of the intervention can be released during the evaluation phase. Second, when evaluating stand-alone interventions in a trial setting, the involvement of HCPs as informants might be crucial as a means of reaching a more diverse group of patients. More than half of the intervention group patients explained that the intervention helped them in the preparation for a clinical consultation, created awareness, and/or reinforced their current communication skills. Patients who reported no benefit explained that they already had a good understanding with their HCP, had a lot of experience, or considered themselves as sufficiently participative and empowered. Finally, it will be interesting to continue monitoring the use and effectiveness of PatientTIME once it is publicly available in order to provide input for other interventions and to keep PatientTIME up-to-date.

## Acknowledgments

The authors would like to thank all participants and Hematon for their kind cooperation, the Alpe d'HuZes program of the Dutch Cancer Society for funding this study (grant number NIVEL 2010-4747), and Jedidja Verboom for her help coding the qualitative data. The funders had no role in this study.

## Conflicts of Interest

None declared.

## References

1. Borrell-Carrió F, Suchman AL, Epstein RM. The biopsychosocial model 25 years later: principles, practice, and scientific inquiry. Ann Fam Med 2004;2(6):576-582 [FREE Full text] [doi: 10.1370/afm.245] [Medline: 15576544]
2. Ong LM, de Haes JC, Hoos AM, Lammes FB. Doctor-patient communication: a review of the literature. Soc Sci Med 1995 Apr;40(7):903-918. [Medline: 7792630]
3. Zolnierek KB, Dimatteo MR. Physician communication and patient adherence to treatment: a meta-analysis. Med Care 2009 Aug;47(8):826-834 [FREE Full text] [doi: 10.1097/MLR.0b013e31819a5acc] [Medline: 19584762]
4. Clever SL, Jin L, Levinson W, Meltzer DO. Does doctor-patient communication affect patient satisfaction with hospital care? Results of an analysis with a novel instrumental variable. Health Serv Res 2008 Oct;43(5 Pt 1):1505-1519 [FREE Full text] [doi: 10.1111/j.1475-6773.2008.00849.x] [Medline: 18459954]
5. Venetis MK, Robinson JD, Turkiewicz KL, Allen M. An evidence base for patient-centered cancer care: a meta-analysis of studies of observed communication between cancer specialists and their patients. Patient Educ Couns 2009 Dec;77(3):379-383. [doi: 10.1016/j.pec.2009.09.015] [Medline: 19836920]
6. Zandbelt LC, Smets EM, Oort FJ, Godfried MH, de Haes HC. Medical specialists' patient-centered communication and patient-reported outcomes. Med Care 2007 Apr;45(4):330-339. [doi: 10.1097/01.mlr.0000250482.07970.5f] [Medline: 17496717]



XSL•FO
RenderX

7. Hack TF, Degner LF, Parker PA. The communication goals and needs of cancer patients: a review. Psychooncology 2005 Oct;14(10):831-45; discussion 846. [doi: 10.1002/pon.949] [Medline: 16200519]
8. Anderson RM, Funnell MM. Patient empowerment: myths and misconceptions. Patient Educ Couns 2010 Jun;79(3):277-282 [FREE Full text] [doi: 10.1016/j.pec.2009.07.025] [Medline: 19682830]
9. Stiggelbout AM, Van der Weijden T, De Wit MP, Frosch D, Légaré F, Montori VM, et al. Shared decision making: really putting patients at the centre of healthcare. BMJ 2012;344:e256 [FREE Full text] [Medline: 22286508]
10. Taylor K. Paternalism, participation and partnership - the evolution of patient centeredness in the consultation. Patient Educ Couns 2009 Feb;74(2):150-155. [doi: 10.1016/j.pec.2008.08.017] [Medline: 18930624]
11. Epstein RM, Franks P, Fiscella K, Shields CG, Meldrum SC, Kravitz RL, et al. Measuring patient-centered communication in patient-physician consultations: theoretical and practical issues. Soc Sci Med 2005 Oct;61(7):1516-1528. [doi: 10.1016/j.socscimed.2005.02.001] [Medline: 16005784]
12. Cegala DJ, Street RL, Clinch CR. The impact of patient participation on physicians' information provision during a primary care medical interview. Health Commun 2007;21(2):177-185. [doi: 10.1080/10410230701307824] [Medline: 17523863]
13. Gattellari M, Butow PN, Tattersall MH. Sharing decisions in cancer care. Soc Sci Med 2001 Jun;52(12):1865-1878. [Medline: 11352412]
14. Degner LF, Kristjanson LJ, Bowman D, Sloan JA, Carriere KC, O'Neil J, et al. Information needs and decisional preferences in women with breast cancer. JAMA 1997 May 14;277(18):1485-1492. [Medline: 9145723]
15. Timmermans LM, van der Maazen RW, Verhaak CM, van Roosmalen MS, van Daal WJ, Kraaimaat FW. Patient participation in discussing palliative radiotherapy. Patient Educ Couns 2005 Apr;57(1):53-61. [doi: 10.1016/j.pec.2004.03.016] [Medline: 15797153]
16. Verhaak CM, Kraaimaat FW, Staps AC, van Daal WA. Informed consent in palliative radiotherapy: participation of patients and proxies in treatment decisions. Patient Educ Couns 2000 Aug;41(1):63-71. [Medline: 10900368]
17. Barg FK, Cronholm PF, Straton JB, Keddem S, Knott K, Grater J, et al. Unmet psychosocial needs of Pennsylvanians with cancer: 1986-2005. Cancer 2007 Aug 1;110(3):631-639 [FREE Full text] [doi: 10.1002/cncr.22820] [Medline: 17592828]
18. Cegala DJ. Emerging trends and future directions in patient communication skills training. Health Commun 2006;20(2):123-129. [doi: 10.1207/s15327027hc2002_3] [Medline: 16965249]
19. Harrington J, Noble LM, Newman SP. Improving patients' communication with doctors: a systematic review of intervention studies. Patient Educ Couns 2004 Jan;52(1):7-16. [Medline: 14729285]
20. van Bruinessen IR, van der Hout LE, van Weel-Baumgarten EM, Gouw H, Zijlstra JM, van DS. Communication during haematological consultations; patients' preferences and professionals' performances. Ann Hematol 2016 Jun;95(7):1177-1183. [doi: 10.1007/s00277-016-2669-z] [Medline: 27091348]
21. Patienttime. 2016. URL: http://www.patienttime.nl/ [accessed 2016-04-14] [WebCite Cache ID 6glBPziUR]
22. van Bruinessen IR, van Weel-Baumgarten EM, Gouw H, Zijlstra JM, Albada A, van DS. Barriers and facilitators to effective communication experienced by patients with malignant lymphoma at all stages after diagnosis. Psychooncology 2013 Dec;22(12):2807-2814. [doi: 10.1002/pon.3352] [Medline: 23897828]
23. Stienen JJ, Ottevanger PB, Wennekes L, van de Schans SA, Dekker HM, Blijlevens NM, et al. Delivering high-quality care to patients with a non-Hodgkin's lymphoma: barriers perceived by patients and physicians. Neth J Med 2014 Jan;72(1):41-48. [Medline: 24457441]
24. van Bruinessen IR, van Weel-Baumgarten EM, Snippe HW, Gouw H, Zijlstra JM, van Dulmen DS. Active patient participation in the development of an online intervention. JMIR Res Protoc 2014;3(4):e59. [doi: 10.2196/resprot.3695] [Medline: 25379679]
25. Maliski SL, Kwan L, Krupski T, Fink A, Orecklin JR, Litwin MS. Confidence in the ability to communicate with physicians among low-income patients with prostate cancer. Urology 2004 Aug;64(2):329-334. [doi: 10.1016/j.urology.2004.03.042] [Medline: 15302489]
26. Maly RC, Frank JC, Marshall GN, DiMatteo MR, Reuben DB. Perceived efficacy in patient-physician interactions (PEPPI): validation of an instrument in older persons. J Am Geriatr Soc 1998 Jul;46(7):889-894. [Medline: 9670878]
27. Campbell M, Fitzpatrick R, Haines A, Kinmonth AL, Sandercock P, Spiegelhalter D, et al. Framework for design and evaluation of complex interventions to improve health. BMJ 2000 Sep 16;321(7262):694-696 [FREE Full text] [Medline: 10987780]
28. Oakley A, Strange V, Bonell C, Allen E, Stephenson J. Process evaluation in randomised controlled trials of complex interventions. BMJ 2006 Feb 18;332(7538):413-416 [FREE Full text] [doi: 10.1136/bmj.332.7538.413] [Medline: 16484270]
29. Saunders RP, Evans MH, Joshi P. Developing a process-evaluation plan for assessing health promotion program implementation: a how-to guide. Health Promot Pract 2005 Apr;6(2):134-147. [doi: 10.1177/1524839904273387] [Medline: 15855283]
30. Hulscher ME, Laurant MG, Grol RP. Process evaluation on quality improvement interventions. Qual Saf Health Care 2003 Feb;12(1):40-46 [FREE Full text] [Medline: 12571344]
31. Moore G, Audrey S, Barker M, Bond L, Bonell C, Hardeman W, et al. Process evaluation of complex interventions: Medical Research Council guidance. BMJ 2015;350:h1258 [FREE Full text] [Medline: 25791983]




XSL•FO
RenderX

32. van Bruinessen IR, van den Ende IT, Visser LN, van Dulmen S. The impact of watching educational video clips on analogue patients' physiological arousal and information recall. Patient Educ Couns 2016 Feb;99(2):243-249. [doi: 10.1016/j.pec.2015.08.022] [Medline: 26427309]
33. Jordan PW. SUS: a “quick and dirty” usability scale. In: Usability evaluation in industry. London: Taylor & Francis; 1996.
34. Zandbelt LC, Smets EM, Oort FJ, Godfried MH, de Haes HC. Satisfaction with the outpatient encounter: a comparison of patients' and physicians' views. J Gen Intern Med 2004 Nov;19(11):1088-1095 [FREE Full text] [doi: 10.1111/j.1525-1497.2004.30420.x] [Medline: 15566437]
35. MAXQDA. 2012. MAXQDA – Professional Software for Qualitative & Mixed Methods Research URL: http://www.maxqda.com/products/maxqda [accessed 2016-07-16] [WebCite Cache ID 6j3GjW0jg]
36. Bangor A, Kortum PT, Miller JT. An Empirical Evaluation of the System Usability Scale. International Journal of Human-Computer Interaction 2008 Jul 30;24(6):574-594. [doi: 10.1080/10447310802205776]
37. Albada A, Ausems MG, Otten R, Bensing JM, van Dulmen S. Use and evaluation of an individually tailored website for counselees prior to breast cancer genetic counseling. J Cancer Educ 2011 Dec;26(4):670-681. [doi: 10.1007/s13187-011-0227-x] [Medline: 21533850]
38. Jones R, Pearson J, McGregor S, Barrett A, Harper GW, Atkinson JM, et al. Does writing a list help cancer patients ask relevant questions? Patient Educ Couns 2002 Aug;47(4):369-371. [Medline: 12135829]
39. Tsulukidze M, Durand MA, Barr PJ, Mead T, Elwyn G. Providing recording of clinical consultation to patients - a highly valued but underutilized intervention: a scoping review. Patient Educ Couns 2014 Jun;95(3):297-304. [doi: 10.1016/j.pec.2014.02.007] [Medline: 24630697]
40. Pitkethly M, Macgillivray S, Ryan R. Recordings or summaries of consultations for people with cancer. Cochrane Database Syst Rev 2008(3):CD001539. [doi: 10.1002/14651858.CD001539.pub2] [Medline: 18646075]
41. Eysenbach G. Consumer health informatics. BMJ 2000 Jun 24;320(7251):1713-1716 [FREE Full text] [Medline: 10864552]
42. Van't RJ, Crutzen R, De Vries H. Investigating predictors of visiting, using, and revisiting an online health-communication program: a longitudinal study. J Med Internet Res 2010;12(3):e37 [FREE Full text] [doi: 10.2196/jmir.1345] [Medline: 20813716]
43. Krijgsman J, Peeters J, Burghouts A, Brabers A, de Jong J, Moll T, et al. Nictiz NIVEL. 2015. Between the spark and the flame: 2015 eHealth monitor URL: http://www.nivel.nl/sites/default/files/bestanden/eHealth_monitor_2015.pdf [accessed 2016-07-18] [WebCite Cache ID 6j5YBbufk]
44. Henselmans I, Heijmans M, Rademakers J, van Dulmen S. Participation of chronic patients in medical consultations: patients' perceived efficacy, barriers and interest in support. Health Expect 2015 Dec;18(6):2375-2388. [doi: 10.1111/hex.12206] [Medline: 24813384]
45. Bensing JM, Tromp F, van Dulmen S, van den Brink-Muinen A, Verheul W, Schellevis FG. Shifts in doctor-patient communication between 1986 and 2002: a study of videotaped general practice consultations with hypertension patients. BMC Fam Pract 2006;7:62 [FREE Full text] [doi: 10.1186/1471-2296-7-62] [Medline: 17064407]
46. Butow P, Devine R, Boyer M, Pendlebury S, Jackson M, Tattersall MH. Cancer consultation preparation package: changing patients but not physicians is not enough. J Clin Oncol 2004 Nov 1;22(21):4401-4409 [FREE Full text] [doi: 10.1200/JCO.2004.66.155] [Medline: 15514382]


## Abbreviations

**HCP:** health care professional
**MRC:** Medical Research Council
**PEPPI:** Perceived Efficacy in Patient-Physician Interactions
**QPS:** question prompt sheet
**RCT:** randomized controlled trial


*Edited by G Eysenbach; submitted 14.04.16; peer-reviewed by A O'Cathain, S Cutrona; comments to author 05.05.16; revised version received 26.05.16; accepted 21.06.16; published 29.07.16.*

*Please cite as:*
*van Bruinessen IR, van Weel-Baumgarten EM, Gouw H, Zijlstra JM, van Dulmen S*
*An Integrated Process and Outcome Evaluation of a Web-Based Communication Tool for Patients With Malignant Lymphoma: Randomized Controlled Trial*
*J Med Internet Res 2016;18(7):e206*
*URL: http://www.jmir.org/2016/7/e206/*
*doi:10.2196/jmir.5877*
*PMID:27473173*




XSL•FO
RenderX

©Inge Renske van Bruinessen, Evelyn M van Weel-Baumgarten, Hans Gouw, Josée M Zijlstra, Sandra van Dulmen. Originally published in the Journal of Medical Internet Research (http://www.jmir.org), 29.07.2016. This is an open-access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/2.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on http://www.jmir.org/, as well as this copyright and license information must be included.



XSL•FO
RenderX

Original Paper

# Paper-Based Survivorship Care Plans May be Less Helpful for Cancer Patients Who Search for Disease-Related Information on the Internet: Results of the Registrationsystem Oncological Gynecology (ROGY) Care Randomized Trial

Kim AH Nicolaije[1,2], PhD; Nicole PM Ezendam[1,2], PhD; Johanna MA Pijnenborg[3], MD, PhD; Dorry Boll[3], MD, PhD; Maria Caroline Vos[3], MD; Roy FPM Kruitwagen[4], MD, PhD; Lonneke V van de Poll-Franse[1,2], PhD

[1]Center of Research on Psychology in Somatic diseases (CoRPS), Department of Medical and Clinical Psychology, Tilburg University, Tilburg, Netherlands

[2]Comprehensive Cancer Center The Netherlands, Eindhoven, Netherlands

[3]Gynecologic Cancer Center South, Department of Gynecology, Elisabeth-TweeSteden Hospital, Tilburg and Waalwijk, Netherlands

[4]Department of Gynecology and GROW, School for Oncology and Developmental Biology, Maastricht University Medical Center, Maastricht, Netherlands

**Corresponding Author:**
Kim AH Nicolaije, PhD
Center of Research on Psychology in Somatic diseases (CoRPS)
Department of Medical and Clinical Psychology
Tilburg University
Warandelaan 2
PO Box 90153
Tilburg, 5000 LE
Netherlands
Phone: 31 134663118
Fax: 31 134662175
Email: k.a.h.schellekens-nicolaije@uu.nl

## Abstract

**Background:** The Institute of Medicine recommends Survivorship Care Plans (SCPs) for all cancer survivors. However, it is unclear whether certain patient groups may or may not benefit from SCPs.

**Objective:** The aim was to assess whether the effects of an automatically generated paper SCP on patients' satisfaction with information provision and care, illness perceptions, and health care utilization were moderated by disease-related Internet use.

**Methods:** Twelve hospitals were randomized to either SCP care or usual care in the pragmatic cluster randomized Registrationsystem Oncological GYnecology (ROGY) Care trial. Newly diagnosed endometrial cancer patients completed questionnaires after diagnosis (N=221; response: 74.7%, 221/296), 6 months (n=158), and 12 months (n=147), including patients' satisfaction with information provision and care, illness perceptions, health care utilization (how many times patients visited a medical specialist or primary care physician about their cancer in the past 6 months), and disease-related Internet use (whether patients used the Internet to look for information about cancer).

**Results:** In total, 80 of 221 (36.2%) patients used the Internet to obtain disease-related information. Disease-related Internet use moderated the SCP care effect on the amount of information received about the disease ($P$=.03) and medical tests ($P$=.01), helpfulness of the information ($P$=.01), and how well patients understood their illness ($P$=.04). All stratified analyses were not statistically significant. However, it appeared that patients who did not seek disease-related information on the Internet in the SCP care arm reported receiving more information about their disease (mean 63.9, SD 20.1 vs mean 58.3, SD 23.7) and medical tests (mean 70.6, SD 23.5 vs mean 64.7, SD 24.9), finding the information more helpful (76.7, SD 22.9 vs mean 67.8, SD 27.2; scale 0-100), and understanding their illness better (mean 6.6, SD 3.0 vs mean 6.1, SD 3.2; scale 1-10) than patients in the usual care arm did. In addition, although all stratified analyses were not significant, patients who did seek disease-related information on the Internet in the SCP care arm appeared to receive less information about their disease (mean 65.7, SD 23.4 vs mean 67.1, SD 20.7) and medical tests (mean 72.4, SD 23.5 vs mean 75.3, SD 21.6), did not find the information more helpful (mean 78.6,



XSL•FO
RenderX

SD 21.2 vs mean 76.0, SD 22.0), and reported less understanding of their illness (mean 6.3, SD 2.8 vs mean 7.1, SD 2.7) than patients in the usual care arm did.

**Conclusions:** Paper SCPs appear to improve the amount of information received about the disease and medical tests, the helpfulness of the information, and understanding of the illness for patients who do not search for disease-related information on the Internet. In contrast, paper SCPs do not seem beneficial for patients who do seek disease-related information on the Internet.

**Trial Registration:** ClinicalTrials.gov NCT01185626; https://clinicaltrials.gov/ct2/show/NCT01185626 (Archived by WebCite at http://www.webcitation.org/6fpaMXsDn)

***(J Med Internet Res 2016;18(7):e162)*** doi:10.2196/jmir.4914

**KEYWORDS**

Survivorship Care Plan; Internet use; pragmatic cluster randomized trial; endometrial neoplasms; patient-reported outcomes; information provision

## Introduction

Information provision has been demonstrated to play an essential role in the support for cancer survivors [1,2]. To improve patient information provision, the Institute of Medicine (IOM) recommends the use of Survivorship Care Plans (SCPs), described as personal treatment summaries and follow-up care plans, for all cancer survivors [3]. However, there is still an ongoing debate about the benefits of SCPs [4-12].

Recent results of the pragmatic cluster randomized Registrationsystem Oncological GYnecology (ROGY) Care trial [8], in which cancer patients were provided with a paper-based SCP, showed that SCPs increased the amount of information received. However, the trial showed no evidence of SCPs benefitting satisfaction with information and care. Furthermore, SCPs increased patients' concerns, emotional impact, experienced symptoms, and the amount of cancer-related contact with the primary care physician. Moreover, it remains unclear whether patient characteristics influence the effects of SCPs and whether certain groups of patients may or may not benefit from SCPs [8].

The SCPs are usually provided by patients' health care providers, who are patients' main source of information about their cancer [1,13]. However, the Internet is also increasingly used as a source of information. Several studies have shown that a significant proportion of cancer survivors, ranging from 30% to 60%, are using the Internet to seek information about their cancer [14-19]. Especially those cancer survivors who are younger [15,17,19], higher educated [15,17,19], male [15], and have a partner [19] use the Internet.

Using the Internet to obtain disease-related information has been associated with considerable benefits for cancer survivors [20]. For instance, it has been found that cancer survivors who use the Internet to access disease-related information feel better informed [15], report receiving more information about their disease and medical tests [21], find the received information more helpful [21], communicate more effectively with their health care providers [22], and are more actively involved in decision making [23]. Therefore, it is possible that receiving an SCP has a different impact on patients who search for information about their cancer on the Internet compared to patients who do not search for information about their cancer on the Internet.

The ROGY Care trial evaluates the impact of an automatically generated SCP on outcomes reported by gynecological cancer patients and health care providers. The trial protocol [24], the primary patient-reported outcomes up to 12 months after diagnosis [8], and the evaluation of the oncology providers [28] and primary care physicians [46] have been previously described. The aim of this analysis of the ROGY Care trial was to assess whether the effects of an automatically generated paper SCP on patients' satisfaction with information provision and care, illness perceptions, and health care utilization were moderated by (ie, different for) disease-related Internet use. It was hypothesized that paper SCPs may be a helpful tool to reach out to patient groups who do not search for information about their cancer on the Internet, whereas SCPs may be of limited value for patients who already benefit from accessing information about their cancer on the Internet.

## Methods

### Design

In the pragmatic cluster randomized controlled ROGY Care trial, 12 hospitals in the Netherlands were randomized to either SCP care or usual care. Patients were included immediately after initial surgery and followed for 24 months. The trial was centrally approved by the Medical Research Ethics Committee of the St Elisabeth Hospital in Tilburg, as well as by the Medical Research Ethics Committees of each participating center [24], and has been registered on ClinicalTrials.gov (NCT01185626). This study describes the results of subgroup analyses of the primary patient-reported outcomes up to 12 months after diagnosis.

### Participants and Recruitment

Participants were women newly diagnosed with endometrial cancer. Exclusion criteria (ie, undergoing palliative care or unable to complete a Dutch questionnaire) [24] were minimal to maximize generalizability [25]. Between April 2011 and October 2012, all eligible patients were invited to participate after initial diagnosis by their own gynecologist by sending a letter, questionnaire, and informed consent form [8,24]. After the first contact through the gynecologist and obtaining informed consent, follow-up questionnaires were sent directly to the home address of the patient at 6 and 12 months after diagnosis .



XSL•FO
RenderX

### Randomization and Blinding

Randomization at the hospital level was chosen to avoid potential contamination of usual care with increased information provision of SCP care and was performed with a table of random numbers by a researcher not involved in the study and blind to the identity of the hospitals. As is common in cluster randomized trials [27], patients were unaware of the assignment to trial arms. Health care providers could not be blinded to trial arm assignment.

### Survivorship Care Plan Versus Usual Care

In the usual care arm, the oncology providers (ie, gynecologists, gynecologic oncologists, oncology nurses) were instructed to continue providing patient information in the way they were used to: they gave standard care according to the Dutch follow-up guidelines, which recommend verbal and written information about the period after treatment and follow-up, signs of recurrence, and hospital contact details. None of the oncology providers in the usual care arm provided SCPs [28].

In the SCP care arm, the oncology providers were instructed to provide an SCP to patients after surgery (ie, during the consultation in which the final histological diagnosis was discussed); to provide an updated SCP during follow-up visits if there were changes in the cancer, treatment, or specialists; and to send a copy of the SCP to the patient's primary care physician. Because of the pragmatic approach of the trial, the delivery of the intervention was allowed to vary between hospitals and oncology providers, fitting their own clinical practice [24].

### Survivorship Care Plan

The Web-based ROGY has been used by all participating oncology providers in both arms since 2006. For each patient, a detailed registration is made in a uniform way, including tumor stage and grade, treatment, comorbidity, complications, follow-up, and information about the involved specialists (eg, gynecologist/gynecologic oncologist, medical oncologist, radiation oncologist). For this trial, an application was built in ROGY enabling automatic generation of an SCP combining patient and disease data by simply pressing a button. The ROGY system was used by all participating oncology providers in both arms, but the SCP button was only visible for oncology providers in the SCP care arm. Any changes related to the cancer, treatment, or specialists were registered in ROGY and automatically updated in the SCP during follow-up.

For the development of the SCP, the Dutch SCP template (based on the IOM format) [3], was adjusted to the local situation [29] by a subgroup of gynecologists/gynecologic oncologists, oncology nurses, a radiation oncologist, a medical oncologist, a primary care physician, and patients [24]. The SCP was pilot-tested on patients with a low/intermediate educational level to ensure that the SCP was understandable.

The SCP consisted of a tailored treatment summary, including information on diagnostic tests, type of cancer, stage, grade, treatment, and contact details of the hospital and specialists. In addition, the SCP contained a tailored follow-up care plan, including detailed information on possible short-term and long-term effects, effects on social and sexual life, possible signs of recurrence and secondary tumors, and information on rehabilitation, psychosocial support, and supportive care services [24].

### Measures

All questionnaires were assessed after initial diagnosis and after 6 and 12 months.

#### *Moderator Variable*

Disease-related Internet use was assessed by asking whether patients had used the Internet to look for information about cancer, which could be answered by either yes or no.

#### *Dependent Variables*

Satisfaction with information provision was assessed with the European Organisation for Research and Treatment of Cancer (EORTC) Quality of Life Group information (QLQ-INFO25) questionnaire [30]. This questionnaire includes four information provision subscales: perceived receipt of information about the disease (four items regarding diagnosis, spread of disease, cause(s) of disease, and whether the disease is under control), medical tests (three items regarding purpose, procedures, and results of tests), treatment (six items regarding medical treatment, benefits, side effects, effects on disease symptoms, social life, and sexual activity) and other care services (four items regarding additional help, rehabilitation options, managing illness at home, psychological support). The question format was as follows: "During your current disease or treatment, how much information have you received on...?" In addition, four single-items were included (information about different places of care, things you can do to help yourself get well, satisfaction with the information, and helpfulness of the information). The answer categories were "not at all," "a little," "quite a bit," and "very much." The scales were converted to 0-100 linear scales, with higher scores indicating better-perceived information provision. Internal consistency for all scales (Cronbach alphas=.70-.87) and test-retest reliability (intraclass correlations=.71-.91) were good [30].

Satisfaction with care was assessed with two multi-item and two single-item scales of the EORTC cancer in-patient satisfaction with care measure (IN-PATSAT32) [31]. This questionnaire was designed to assess cancer patients' perception of the quality of medical care, nursing care, and care organization and services received in the hospital. The multi-item scales included doctors' and nurses' interpersonal skills. The single-item scales included exchange of information between caregivers and general satisfaction with care. The question format was as follows: "How would you rate...?" The answer categories were "poor," "fair," "good," "very good," and "excellent." The scales were converted to 0-100 linear scales, with higher scores indicating better-perceived quality of care. Internal consistency (Cronbach alphas=.67-.96) and test-retest reliability (intraclass correlations=.66-.85) were good [31].

Illness perception was assessed with the Brief Illness Perception Questionnaire (B-IPQ) [32], consisting of eight single-item scales, measuring cognitive representations (consequences,



XSL•FO
RenderX

timeline, personal control, treatment control, identity), emotional representations (concern, emotion), and illness comprehensibility rated on a 0-10 linear scale, with higher scores indicating more endorsement of that item. Test-retest reliability (Pearson correlations=.42-.75) was good [32].

Health care utilization was assessed by asking how many times patients visited a medical specialist or primary care physician in relation to cancer in the past 6 months. These questions were asked in a similar way as is done by Statistics Netherlands.

### *Control Variables*

Sociodemographic and clinical information were obtained from ROGY (ie, date of birth, date of diagnosis, disease stage, primary treatment) and the questionnaire (ie, marital status, educational level as an indicator for socioeconomic status [SES], employment status). Comorbidity was assessed by the adapted Self-administered Comorbidity Questionnaire (SCQ) [33].

### Statistical Analyses

Statistical analyses were conducted using SPSS version 19.0 (IBM Corp, Armonk, NY, USA). Tests were two-sided and considered significant if $P<.05$. Both intention-to-treat and per protocol analyses were conducted. Intention-to-treat analyses compared all respondents in the SCP care arm to all respondents in the usual care arm. Per protocol analyses compared respondents in the SCP care arm who indicated receiving an SCP in the first questionnaire to all respondents in the usual care arm. Because intention-to-treat and per protocol analyses revealed similar results, only the results of the intention-to-treat analyses are reported in this study.

Means with standard deviations were used to describe continuous variables and frequencies with percentages to describe categorical variables. Differences in sociodemographic and clinical characteristics between respondents and nonrespondents, between the SCP care arm and the usual care arm, and between patients who did or did not use the Internet to obtain information about their disease were compared using *t* tests for continuous variables and chi-square tests for categorical variables.

Moderation of disease-related Internet use on the dependent variables (ie, 22 scales in total: eight on information provision, four on satisfaction with care, eight on illness perceptions, and two on health care utilization) was tested by assessing the significance of the interaction term "trial arm × disease-related Internet use" in the overall linear multilevel regression model. Multilevel analysis corrects for missing data (assumed missing at random) by using information from the observed outcomes to provide information about the unobserved outcomes [34,35] The model included two random intercepts (ie, hospital- and patient-level) to account for both clustering at hospital-level and intrapatient dependency of repeated measures [36], the independent variables intervention arm (ie, SCP care vs usual care) and time, the covariates age, time since diagnosis, marital status, employment, educational level, comorbidity, disease stage, and treatment, and the dependent variables information provision and care, illness perceptions, and health care utilization. For the models that did not converge, hospital was included as covariate instead of as random intercept [37].

When an interaction term was significant, this was an indication that the effect of providing an SCP was different for patients who did or did not use the Internet to search for disease-related information and that stratified analyses were warranted to further explore the direction of the moderation effects. For significant interaction terms, the intervention effects were re-examined in subgroups by performing the overall linear multilevel regression analyses stratified by the levels of the moderator variable (ie, disease-related Internet use). Unstandardized betas were presented with 95% confidence intervals.

The trial was originally powered to detect a clinically meaningful difference on the overall primary outcomes of the intervention, targeting 75 patients per arm [8,24]. The trial was not powered to detect differences in moderation analyses or stratified analyses. In this study, moderation analyses and stratified analyses were performed despite this lack of power because we merely wanted to explore the potential moderating role of Internet use. These analyses can be justified because they are exploratory and because the exploration was a priori restricted to a selected moderator with a specific rationale [38].

## Results

### Patient Characteristics

Of the 296 eligible patients, 221 (74.7%) patients completed the first questionnaire. After 6 months, 158 patients completed the questionnaire; after 12 months, 147 patients completed the questionnaire (Figure 1) [8].

At baseline, participants were younger (mean 67.4, SD 8.9 years) than nonparticipants (mean 70.2, SD 9.5 years, $P=.02$), and more often had an International Federation of Gynecology and Obstetrics (FIGO) staging level of stage I (85.5%, 189/221 vs 69%, 52/75; $P=.003$; Table 1) [8]. In total, 80 of 221 (36.2%) patients indicated that they used the Internet to obtain information about their disease. This did not differ between the SCP care arm and the usual care arm (Table 2).



XSL•FO
RenderX

**Table 1.** CONSORT table of baseline sociodemographic and clinical characteristics of participants according to trial arm and of nonparticipants.

| Patient characteristics | | SCP care (n=119) | Usual care (n=102) | *P* [a] | Total participants (N=221) | Nonparticipants (n=75) | *P* [a] |
|---|---|---|---|---|---|---|---|
| Age at diagnosis, mean (SD) | | 67.1 (9.1) | 67.7 (8.8) | .65 | 67.4 (8.9) | 70.2 (9.5) | .02 |
| **FIGO stage, n (%)** | | | | | | | |
| | I | 102 (85.7) | 87 (85.3) | .75 | 189 (85.5) | 52 (69) | .003 |
| | II | 5 (4.2) | 2 (2.0) | | 7 (3.2) | 10 (13) | |
| | II | 8 (6.7) | 7 (6.8) | | 15 (6.8) | 11 (15) | |
| | IV | 3 (2.9) | 4 (3.9) | | 7 (3.2) | 1 (1) | |
| **Treatment, n (%)** | | | | | | | |
| | Surgery | 117 (98.3) | 97 (95) | .46 | 214 (96.8) | 72 (96) | .45 |
| | Radiotherapy | 44 (37.0) | 37 (36.3) | .99 | 81 (36.7) | 34 (45) | .19 |
| | Chemotherapy | 6 (5.0) | 12 (11.8) | .06 | 18 (8.1) | 7 (9) | .76 |
| **Hospital, n (%)** | | | | | | | |
| | 1 | 22 (18.5) | | | 22 (10.0) | 4 (5) | .61 |
| | 2 | 12 (10.1) | | | 12 (5.4) | 7 (9) | |
| | 3 | 28 (23.5) | | | 28 (12.7) | 9 (12) | |
| | 4 | 28 (23.5) | | | 28 (12.7) | 9 (12) | |
| | 5 | 11 (9.2) | | | 11 (5.0) | 1 (1) | |
| | 6 | 18 (15.1) | | | 18 (8.1) | 5 (7) | |
| | 7 | | 25 (24.5) | | 25 (11.3) | 13 (17) | |
| | 8 | | 21 (20.5) | | 21 (9.5) | 6 (8) | |
| | 9 | | 26 (25.5) | | 26 (11.8) | 7 (9) | |
| | 10 | | 12 (11.8) | | 12 (5.4) | 4 (5) | |
| | 11 | | 3 (2.9) | | 3 (1.4) | 3 (4) | |
| | 12 | | 15 (14.7) | | 15 (6.8) | 7 (9) | |

[a]*P* values report comparisons between the intervention arm and the usual care arm, and between the trial participants and nonparticipants according to *t* tests and chi-square tests.



XSL•FO
RenderX

**Table 2.** Sociodemographic and clinical characteristics at the first questionnaire according to trial arm.

| Patient characteristics | SCP care (n=119) | Usual care (n=102) | *P* [a] | Total (N=221) |
|---|---|---|---|---|
| Age at time of survey, mean (SD) | 67.4 (9.1) | 67.8 (8.9) | .71 | 67.6 (9.0) |
| Months since diagnosis, mean (SD) | 2.6 (1.7) | 1.8 (1.2) | <.001 | 2.1 (1.5) |
| **Months since diagnosis, n (%)** | | | | |
| <1 | 12 (10.1) | 24 (23.5) | | 36 (16.3) |
| 1-2 | 40 (33.6) | 46 (45.0) | | 86 (38.9) |
| 2-3 | 33 (27.7) | 20 (19.6) | | 53 (24.0) |
| >3 | 34 (28.6) | 12 (11.8) | | 46 (20.8) |
| **Comorbidity, n (%)** | | | .41 | |
| None | 19 (16.0) | 18 (17.6) | | 37 (16.7) |
| 1 | 32 (26.9) | 20 (19.6) | | 52 (23.5) |
| ≥2 | 64 (53.8) | 62 (60.8) | | 126 (57.0) |
| **Marital status, [b] n (%)** | | | .74 | |
| Partner | 85 (71.4) | 76 (74.5) | | 161 (72.9) |
| No partner | 31 (26.1) | 25 (24.5) | | 56 (25.3) |
| **Educational level,[c]n (%)** | | | .09 | |
| High | 17 (14.3) | 7 (6.9) | | 24 (10.9) |
| Intermediate | 71 (59.7) | 72 (70.6) | | 143 (64.7) |
| Low | 30 (25.2) | 19 (18.6) | | 49 (22.2) |
| **Employed, n (%)** | | | .40 | |
| Yes | 22 (18.5) | 15 (14.7) | | 37 (16.7) |
| No | 85 (71.4) | 79 (77.5) | | 164 (74.2) |
| **Disease-related Internet use, n (%)** | | | .57 | |
| Yes | 41 (34.5) | 39 (38.2) | | 80 (36.2) |
| No | 76 (63.9) | 60 (58.8) | | 136 (61.5) |

[a]*P* values report comparisons between the intervention arm and the usual care arm, according to *t* tests and chi-square tests.

[b]Marital status: partner=married/living together, no partner=divorced/widowed/never married.

[c]Educational level: low=no/primary school, intermediate=lower general secondary education/vocational training, high=preuniversity education/ high vocational training/university.

Patients who used the Internet to obtain disease-related information were younger (mean 62.8, SD 7.5 years) than patients who did not use the Internet to obtain disease-related information (mean 70.3, SD 8.7 years; Table 3). In addition, patients who used the Internet to obtain disease-related information more often had a partner (83%, 66/80 vs 67.6%, 92/136), more often had a high educational level (20%, 16/80 vs 5.1%, 7/136), and were employed more often (28%, 21/80 vs 11.8%, 16/136) than patients who did not use the Internet to obtain disease-related information.



XSL•FO
RenderX

**Table 3.** Patient characteristics at the first questionnaire according to disease-related Internet use.

| Patient characteristics | Disease-related Internet use (n=80) | No disease-related Internet use (n=136) | *P* [a] | Total (N=216) |
|---|---|---|---|---|
| Age at time of survey, mean (SD) | 62.8 (7.5) | 70.3 (8.7) | <.001 | 67.5 (9.0) |
| Months since diagnosis, mean (SD) | 2.0 (1.3) | 2.3 (1.6) | .10 | 2.2 (1.5) |
| **Months since diagnosis, n (%)** | | | | |
| <1 | 19 (24) | 16 (11.8) | | 35 (16.2) |
| 1-2 | 31 (39) | 53 (39.0) | | 84 (38.9) |
| 2-3 | 12 (15) | 41 (30.1) | | 53 (24.5) |
| >3 | 18 (23) | 26 (19.1) | | 44 (20.4) |
| **FIGO stage, n (%)** | | | .37 | |
| I | 70 (88) | 115 (84.6) | | 185 (85.6) |
| II | 4 (5) | 3 (2.2) | | 7 (3.2) |
| II | 3 (4) | 12 (8.8) | | 15 (6.9) |
| IV | 2 (3) | 4 (2.9) | | 6 (2.8) |
| **Treatment, n (%)** | | | | |
| Surgery | 77 (96) | 132 (97.1) | .90 | 209 (96.8) |
| Radiotherapy | 26 (33) | 54 (39.7) | .31 | 80 (37.0) |
| Chemotherapy | 7 (9) | 10 (7.4) | .70 | 17 (7.9) |
| **Comorbidity, n (%)** | | | .20 | |
| None | 18 (23) | 19 (14.0) | | 37 (17.1) |
| 1 | 15 (19) | 35 (25.7) | | 50 (23.1) |
| ≥2 | 45 (56) | 78 (57.4) | | 123 (56.9) |
| **Marital status, [b] n (%)** | | | .01 | |
| Partner | 66 (83) | 92 (67.6) | | 158 (73.1) |
| No partner | 12 (15) | 42 (30.9) | | 54 (25.0) |
| **Educational level, [c] n (%)** | | | <.001 | |
| High | 16 (20) | 7 (5.1) | | 23 (10.6) |
| Intermediate | 59 (74) | 83 (61.0) | | 142 (65.7) |
| Low | 5 (6) | 42 (30.9) | | 47 (21.8) |
| **Employed, n (%)** | | | .01 | |
| Yes | 21 (28) | 16 (11.8) | | 37 (17.1) |
| No | 54 (72) | 107 (78.7) | | 161 (74.5) |

[a]*P* values report comparisons between patients reporting disease-related Internet use and patients not reporting disease-related Internet use according to *t* tests and chi-square tests.

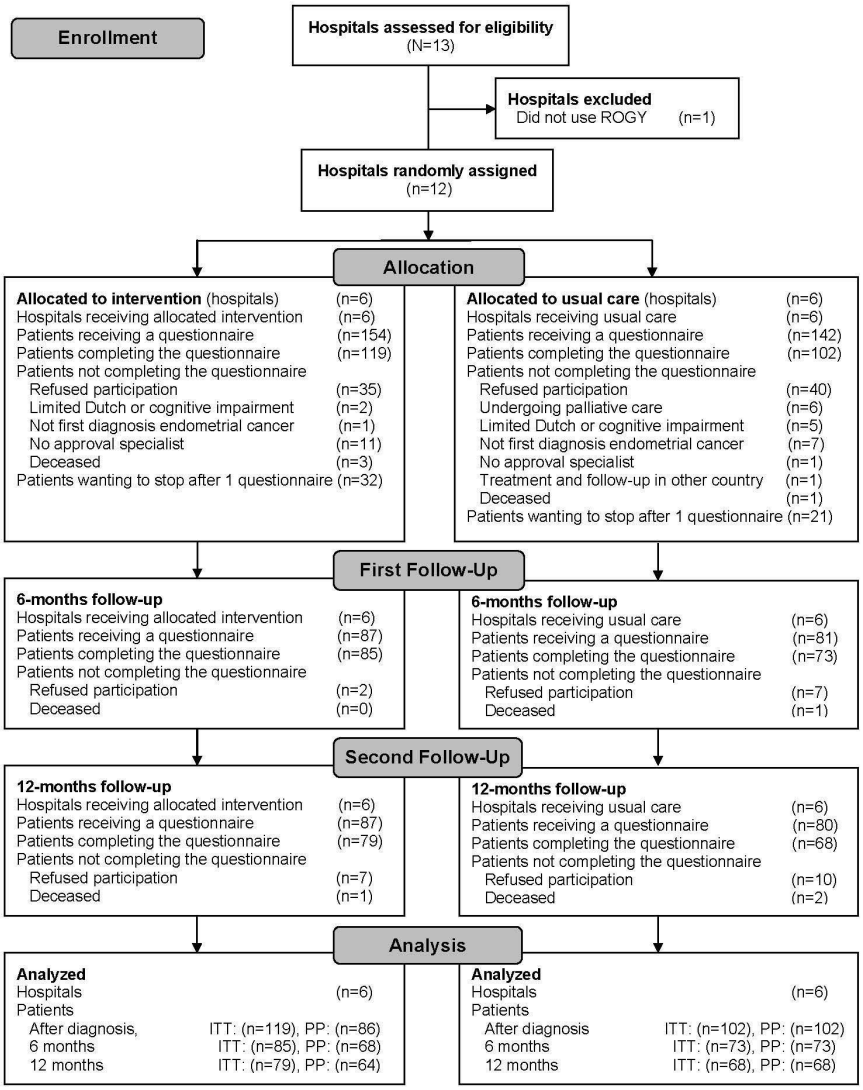
[b]Marital status: partner=married/living together, no partner=divorced/widowed/never married.

[c]Educational level: low=no/primary school, intermediate=lower general secondary education/vocational training, high=preuniversity education/high vocational training/university.



XSL•FO
RenderX

**Figure 1.** CONSORT flow diagram of the progress of the hospitals and endometrial cancer patients through the phases of the ROGY Care trial. ITT: intention-to-treat analyses (comparing all respondents in the SCP care arm to all respondents in the usual care arm); PP: per protocol analyses (comparing the respondents in the SCP care arm who indicated that they received an SCP in the first questionnaire to all respondents in the usual care arm).




## Moderation of Disease-Related Internet Use

Four statistically significant moderation tests were found. Disease-related Internet use moderated the intervention effect on the amount of information received about the disease ($P$=.03), the amount of information received about medical tests ($P$=.01), the helpfulness of the information ($P$=.01), and how well patients understand their illness ($P$=.04). All other interaction terms were not significant.

Although the stratified analyses were all not statistically significant, it appeared that patients who did not seek disease-related information on the Internet may have benefitted from receiving an SCP because patients in the SCP care arm reported receiving more information about their disease (mean 63.9, SD 20.1 vs mean 58.3, SD 23.7) and medical tests (mean 70.6, SD 23.5 vs mean 64.7, SD 24.9), found the information more helpful (mean 76.7, SD 22.9 vs mean 67.8, SD 27.2), and understood their illness better (mean 6.6, SD 3.0 vs mean 6.1, SD 3.2) than patients in the usual care arm did (Table 4 and Figures 2-5). On the other hand, although the stratified analyses were all not statistically significant, it appeared that patients who did seek disease-related information on the Internet did not benefit from receiving an SCP because patients in the SCP care arm did not report receiving more information about their disease (mean 65.7, SD 23.4 vs mean 67.1, SD 20.7) and medical tests (mean 72.4, SD 23.5 vs mean 75.3, SD 21.6), did not find the information more helpful (mean 78.6, SD 21.2 vs mean 76.0, SD 22.0), and reported less understanding of their illness (mean 6.3, SD 2.8 vs mean 7.1, SD 2.7) than patients in the usual care arm did (Table 4 and Figures 2-5).



XSL•FO
RenderX

**Table 4.** Regression outcomes from the stratified analyses for the effect of SCP care on the outcomes according to disease-related Internet use.

| Outcome[a] | SCP care, mean (SD)[b] | Usual care, mean (SD)[b] | Total, mean (SD) | Beta (95% CI)[c] | *P* |
|---|---|---|---|---|---|
| **Information disease** [d] | | | | | |
| Internet use | 65.7 (23.4) | 67.1 (20.7) | 66.4 (22.1) | –1.36 (–12.7, 10.0) | .79 |
| No Internet use | 63.9 (20.1) | 58.3 (23.7) | 61.4 (21.9) | 5.51 (–3.9, 14.9) | .22 |
| **Information medical tests** [d] | | | | | |
| Internet use | 72.4 (23.5) | 75.3 (21.6) | 73.9 (22.6) | –3.83 (–13.5, 5.8) | .43 |
| No Internet use | 70.6 (23.5) | 64.7 (24.9) | 68.0 (24.3) | 4.87 (–3.3, 13.0) | .24 |
| **Helpfulness information** [d] | | | | | |
| Internet use | 78.6 (21.2) | 76.0 (22.0) | 77.3 (21.6) | 1.13 (–7.4, 9.6) | .79 |
| No Internet use | 76.7 (22.9) | 67.8 (27.2) | 72.9 (25.2) | 6.89 (–1.6, 15.4) | .11 |
| **How well understand illness** [e] | | | | | |
| Internet use | 6.3 (2.8) | 7.1 (2.7) | 6.7 (2.8) | –0.98 (–2.11, 0.14) | .09 |
| No Internet use | 6.6 (3.0) | 6.1 (3.2) | 6.3 (3.1) | 0.30 (–0.73, 1.33) | .56 |

[a]Outcomes are presented only for the statistically significant interaction terms. Linear multilevel regression analyses were performed, adjusted for age, time since diagnosis, marital status, educational level, employment, comorbidities, stage, and treatment. For the models that did not converge, hospital was included as covariate instead of random intercept.

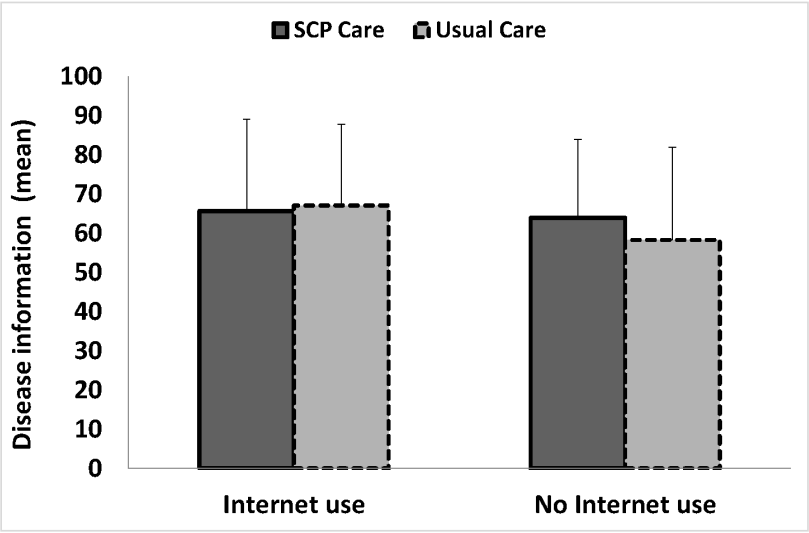
[b]Crude means and standard deviations are reported for SCP care and usual care.

[c]Unstandardized betas and 95% confidence intervals are reported for SCP care (ref=usual care).

[d]EORTC-QLQ-INFO25 scale range from 0-100: higher scores reflect better-perceived information received.

[e]B-IPQ scale range from 1-10: higher scores indicate more endorsement of that item.

**Figure 2.** Patients' reported amount of information received about their disease according to disease-related Internet use for the SCP care and the usual care arms. Crude means are reported. Error bars represent +1 SD. EORTC-QLQ-INFO25 scale ranges from 0-100 (higher scores reflect better perceived information received).






XSL•FO
RenderX

**Figure 3.** Patients' reported amount of information received about their medical tests according to disease-related Internet use for the SCP care and the usual care arms. Crude means are reported. Error bars represent +1 SD. EORTC-QLQ-INFO25 scale ranges from 0-100 (higher scores reflect better perceived information received).

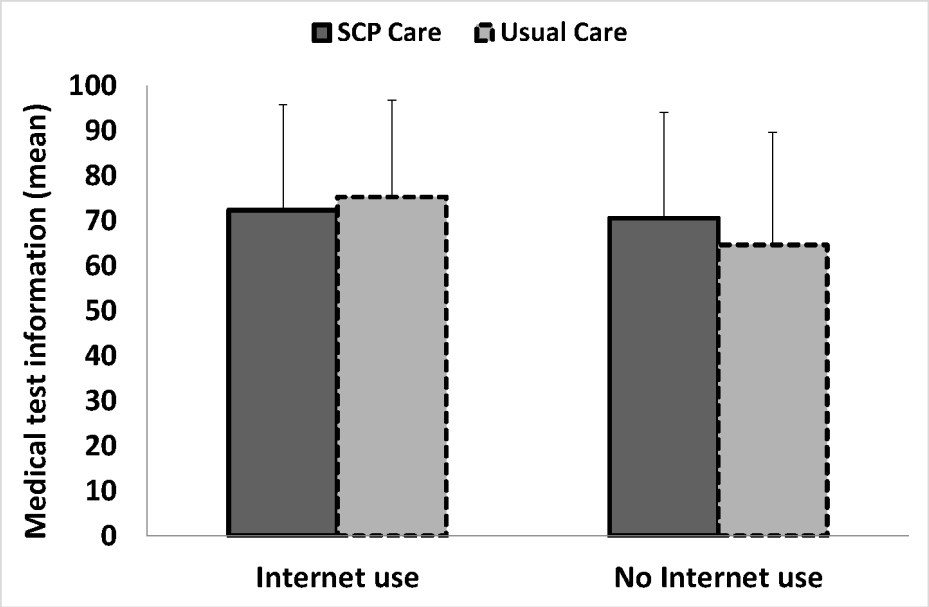
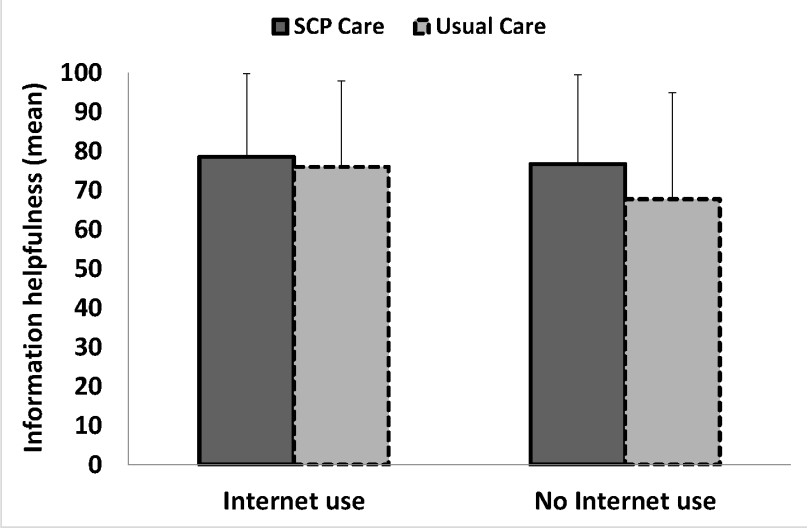


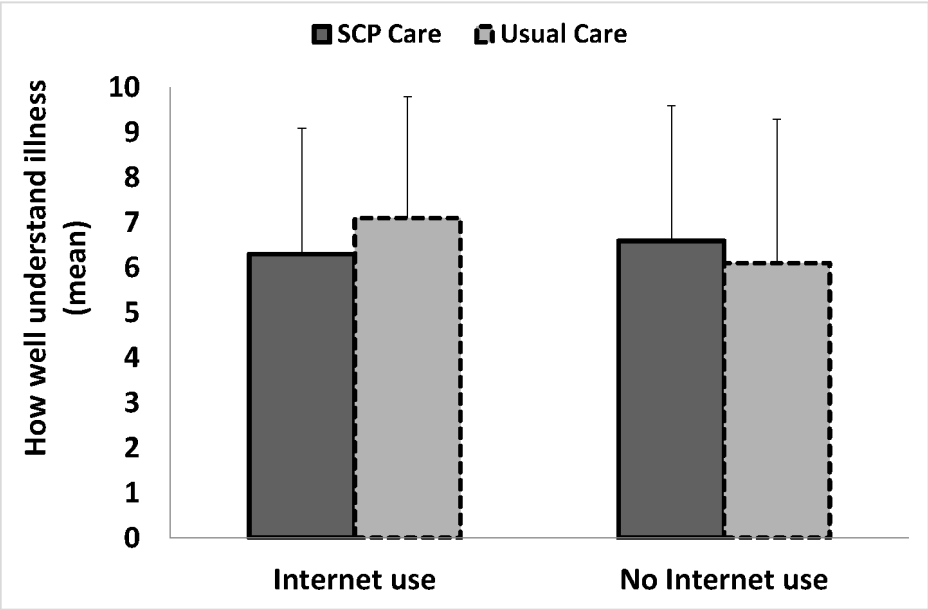

**Figure 4.** Patient-reported helpfulness of the information received according to disease-related Internet use for the SCP care arm and the usual care arm. Crude means are reported. Error bars represent +1 SD. EORTC-QLQ-INFO25 scale ranges from 0-100 (higher scores reflect better-perceived information received).






XSL•FO
RenderX

**Figure 5.** Patients' reported understanding of their illness according to disease-related Internet use for the SCP care and the usual care arms. Crude means are reported. Error bars represent +1 SD. B-IPQ scale ranges from 1-10 (higher scores indicate more endorsement of that item).




## Discussion

The results of this secondary analysis of the ROGY Care trial suggest that paper-based SCPs appear to improve the amount of received information about the disease and medical tests, the helpfulness of the information, and the understanding of the illness for patients who do not search for information on the Internet themselves. In contrast, paper-based SCPs do not appear helpful for patients who already search for information on the Internet themselves. All other outcomes did not differ for patients who did or did not use the Internet to search for disease-related information.

### Patients Who Did Not Use the Internet to Search for Disease-Related Information

Nearly two-thirds of all patients in this study did not use the Internet to search for disease-related information. These patients were older, lower educated, and less often had a partner or a job than patients who did use the Internet to search for information about their cancer. This has consistently been found in previous studies [15,17,19] and has raised the concern that some patient groups do not equally benefit from the various resources available on the Internet [19]. Because educational level is an indicator for SES [39,40], patients with a higher SES search the Internet more for disease-related information than patients with a lower SES. This "digital divide" may pose a threat to equity in health care when important information can only be or best be accessed online [19]. Even today, a large number of cancer survivors do not have access to the potential benefits of the Internet. The results of this study suggest that paper-based SCPs may be a useful tool to empower this patient group by increasing the amount of information they receive about their disease and medical tests, the helpfulness of the information, and their understanding of their illness.

### Patients Who Did Use the Internet to Search for Disease-Related Information

A third of all patients in this study did use the Internet to search for disease-related information, which is consistent with previous studies [14-19]. The results of this study suggest that paper-based SCPs may not be of added value for this patient group. A possible explanation for this finding could be that these patients already benefit from accessing information on the Internet because using the Internet to obtain disease-related information has been associated with considerable benefits for cancer survivors [20]. Previous studies have found that cancer survivors who use the Internet to access disease-related information feel better informed [15], report receiving more information about their disease and medical tests [21], find the received information more helpful [21], communicate more effectively with their health care providers [22], and are more actively involved in decision making [23].

Surprisingly, the results of this study suggest that paper-based SCPs may actually even decrease patients' understanding of their illness for those patients who search for disease-related information on the Internet. A possible explanation could be that patients who receive an SCP and also search for information on the Internet may find information on the Internet about their illness that conflicts with information within the SCP. This may confuse patients and may lower their understanding of the illness. Because these patients have access to more information, they may also be more aware of aspects of their illness that they do not (completely) understand (ie, the more you know, the more you realize how little you know). Future research needs



XSL•FO
RenderX

to investigate why SCPs may not be helpful for patients who search for disease-related information on the Internet. Another possibility is that paper-based SCPs in their current form are not suitable for patients who search for disease-related information on the Internet. A possible way to increase the value of SCP care for patients who search for disease-related information on the Internet may be to provide these patients with access to a tailored online SCP instead of a paper-based SCP. Previous research showed that most patients who use the Internet prefer to get their information from reliable websites, such as their hospital's website, and would like to have online access to their own medical file and test results [15]. Internet-based SCPs may be a useful way to support these patients in finding reliable information online that is tailored to their specific situation. The results of previous studies investigating cancer patients' satisfaction with an Internet-based SCP tool seem promising [41-43]. Future research needs to examine whether dissemination of tailored online SCPs does have added value for patients who search for cancer-related information on the Internet.

## Considerations

It is important to take into consideration that this study was conducted in the Netherlands, a developed country where 95% of the population has access to the Internet at home [44]. Furthermore, only endometrial cancer patients were included in this study. In general, endometrial cancer patients have a lower educational level than patients with other types of cancer [45] and lower educational levels have been found to be strongly associated with lower Internet use [15,17,19]. In addition, men tend to use the Internet more often than women [15]. Consequently, the percentage of patients who used the Internet to search for disease-related information in this study may be an underestimation of the Internet use of cancer survivors in the Netherlands. A previous study conducted in the Netherlands in 2006 that included both male and female patients with different types of cancer found that 60% reported using the Internet by themselves [15].

Other effects of SCP care found in the ROGY Care trial [8], such as increased concerns about the illness, emotional impact, experienced symptoms, and health care utilization, did not differ for patients who did or did not use the Internet to search for disease-related information. This finding indicates that SCPs increase patients' concerns, emotional impact, experienced symptoms, and health care utilization for both patients who do and do not search for disease-related information on the Internet. It is possible that certain aspects of the SCP that are not found on the Internet (eg, receiving information from the physician, receiving personalized information, and receiving information about additional care) lead to increased concerns, emotional impact, experienced symptoms, and health care utilization. However, it is important to consider that the ROGY Care trial was not originally powered to detect differences in moderation analyses or stratified analyses. Therefore, it is unclear whether insignificant outcomes in these analyses indicate that disease-related Internet use did not moderate these outcomes or that the power was merely too small to find the effects. On the other hand, this does make the moderation effects that were found in this study more convincing.

## Strengths and Limitations

A limitation of this study is that self-reported information provision and health care utilization were assessed, which makes it unclear how much information was actually provided and how much health care was actually used. In addition, Internet utilization was measured with a single dichotomous item. Consequently, this study can only make a distinction between patients who did or did not use the Internet to search for disease-related information. For instance, it remains unknown how many times patients searched the Internet, what they searched for (ie, did they use the Internet to search for similar topics as addressed in the SCP?), or what information they found. For future research, we recommend using a more elaborate measure of Internet utilization that is psychometrically tested.

Despite these limitations, this study provides important new insight into whether certain groups of patients may or may not benefit from paper-based SCPs in routine clinical practice. The pragmatic cluster randomized design, limited exclusion criteria, and high response rate improve the generalizability of the findings. However, there is not enough evidence to recommend that patients who search for information on the Internet should not receive a paper-based SCP. More research is needed to get a more nuanced understanding of these findings before health care providers can use the information to decide whether providing a paper-based SCP is of added value or not. In addition, future research needs to examine whether other patient characteristics could also possibly influence the impact of SCPs.

## Conclusions

The results of this secondary analyses of the ROGY Care trial suggest that paper-based SCPs may improve the amount of received information about the disease and medical tests, the helpfulness of the information, and the understanding of the illness for patients who do not search for information on the Internet themselves. In contrast, paper-based SCPs do not seem beneficial for patients who do search for disease-related information on the Internet. With the increasing importance of the Internet as a source of information for cancer survivors, future research needs to examine whether dissemination of tailored online SCPs may have added value for patients who use the Internet to obtain disease-related information.

## Acknowledgments

The ROGY Care trial was supported by grant no. UVT 2010-4743 from the Dutch Cancer Society. Prof Lonneke van de Poll-Franse was supported by a Cancer Research Award from the Dutch Cancer Society (#UVT-2009-4349). The Dutch Cancer Society did not contribute to the design and conduct of the study; collection, management, analysis, and interpretation of the data; preparation, review, or approval of the manuscript; or decision to submit the manuscript for publication.



XSL•FO
RenderX

The authors would like to thank all patients and health care providers for their participation in the study. Special thanks go to the research assistants, in particular Mrs E de Winter, who helped to collect the data for this study. In addition, we would like to acknowledge Dr M van Bommel, who was willing to function as an independent advisor and to answer patients' questions. We would like to thank the following hospitals and institutions for their cooperation: Amphia Hospital, Breda and Oosterhout; Catharina Hospital, Eindhoven; Elisabeth-TweeSteden Hospital, Tilburg and Waalwijk; Elkerliek Hospital, Helmond; Jeroen Bosch Hospital's-Hertogenbosch; Institute Verbeeten, Tilburg; Maastricht University Medical Center, Maastricht; Maxima Medical Center, Eindhoven and Veldhoven; Orbis Medical Center, Sittard; St. Anna Hospital, Geldrop and Eindhoven; St. Jans Hospital, Weert; and VieCuri Hospital, Venlo and Venray.

## Conflicts of Interest

None declared.

## Multimedia Appendix 1

Example of the ROGY Care SCP.

[PDF File (Adobe PDF File), 1MB - jmir_v18i7e162_app1.pdf ]

## References

1. Rutten LJ, Arora NK, Bakos AD, Aziz N, Rowland J. Information needs and sources of information among cancer patients: a systematic review of research (1980-2003). Patient Educ Couns 2005 Jun;57(3):250-261. [doi: 10.1016/j.pec.2004.06.006] [Medline: 15893206]
2. Husson O, Mols F, van de Poll-Franse LV. The relation between information provision and health-related quality of life, anxiety and depression among cancer survivors: a systematic review. Ann Oncol 2011 Apr;22(4):761-772 [FREE Full text] [doi: 10.1093/annonc/mdq413] [Medline: 20870912]
3. Hewitt M, Greenfield S, Stovall E, editors. From Cancer Patient to Cancer Survivor: Lost in Transition. Washington, DC: National Academies Press; 2006.
4. Mayer DK, Birken SA, Check DK, Chen RC. Summing it up: an integrative review of studies of cancer survivorship care plans (2006-2013). Cancer 2015 Apr 1;121(7):978-996. [doi: 10.1002/cncr.28884] [Medline: 25252164]
5. Grunfeld E, Julian JA, Pond G, Maunsell E, Coyle D, Folkes A, et al. Evaluating survivorship care plans: results of a randomized, clinical trial of patients with breast cancer. J Clin Oncol 2011 Dec 20;29(36):4755-4762 [FREE Full text] [doi: 10.1200/JCO.2011.36.8373] [Medline: 22042959]
6. Brothers BM, Easley A, Salani R, Andersen BL. Do survivorship care plans impact patients' evaluations of care? A randomized evaluation with gynecologic oncology patients. Gynecol Oncol 2013 Jun;129(3):554-558 [FREE Full text] [doi: 10.1016/j.ygyno.2013.02.037] [Medline: 23474344]
7. Hershman DL, Greenlee H, Awad D, Kalinsky K, Maurer M, Kranwinkel G, et al. Randomized controlled trial of a clinic-based survivorship intervention following adjuvant therapy in breast cancer survivors. Breast Cancer Res Treat 2013 Apr;138(3):795-806. [doi: 10.1007/s10549-013-2486-1] [Medline: 23542954]
8. Nicolaije KA, Ezendam NP, Vos M, Pijnenborg JM, Boll D, Boss EA, et al. Impact of an automatically generated cancer survivorship care plan on patient-reported outcomes in routine clinical practice: longitudinal outcomes of a pragmatic, cluster randomized trial. J Clin Oncol 2015 Nov 1;33(31):3550-3559. [doi: 10.1200/JCO.2014.60.3399] [Medline: 26304900]
9. Smith TJ, Snyder C. Is it time for (survivorship care) plan B? J Clin Oncol 2011 Dec 20;29(36):4740-4742 [FREE Full text] [doi: 10.1200/JCO.2011.38.8397] [Medline: 22042961]
10. Jefford M, Schofield P, Emery J. Improving survivorship care. J Clin Oncol 2012 Apr 20;30(12):1391-1392; author reply 1393 [FREE Full text] [doi: 10.1200/JCO.2011.40.5886] [Medline: 22291077]
11. Stricker CT, Jacobs LA, Palmer SC. Survivorship care plans: an argument for evidence over common sense. J Clin Oncol 2012 Apr 20;30(12):1392-1393; author reply 1393 [FREE Full text] [doi: 10.1200/JCO.2011.40.7940] [Medline: 22291072]
12. Keesing S, McNamara B, Rosenwax L. Cancer survivors' experiences of using survivorship care plans: a systematic review of qualitative studies. J Cancer Surviv 2015 Jun;9(2):260-262 [FREE Full text] [doi: 10.1007/s11764-014-0407-x] [Medline: 25343971]
13. Newnham GM, Burns WI, Snyder RD, Dowling AJ, Ranieri NF, Gray EL, et al. Information from the Internet: attitudes of Australian oncology patients. Intern Med J 2006 Nov;36(11):718-723. [doi: 10.1111/j.1445-5994.2006.01212.x] [Medline: 17040358]
14. Helft PR, Hlubocky F, Daugherty CK. American oncologists' views of internet use by cancer patients: a mail survey of American Society of Clinical Oncology members. J Clin Oncol 2003 Mar 1;21(5):942-947. [Medline: 12610198]
15. van de Poll-Franse LV, van Eenbergen MC. Internet use by cancer survivors: current use and future wishes. Support Care Cancer 2008 Oct;16(10):1189-1195. [doi: 10.1007/s00520-008-0419-z] [Medline: 18293014]
16. Chen X, Siu LL. Impact of the media and the internet on oncology: survey of cancer patients and oncologists in Canada. J Clin Oncol 2001 Dec 1;19(23):4291-4297. [Medline: 11731511]



XSL•FO
RenderX

17. Smith RP, Devine P, Jones H, DeNittis A, Whittington R, Metz JM. Internet use by patients with prostate cancer undergoing radiotherapy. Urology 2003 Aug;62(2):273-277. [Medline: 12893334]
18. Fogel J, Albert S, Schnabel F. Use of the internet by women with breast cancer. J Med Internet Res 2002;4(2):e9-e9.
19. Kowalski C, Kahana E, Kuhr K, Ansmann L, Pfaff H. Changes over time in the utilization of disease-related Internet information in newly diagnosed breast cancer patients 2007 to 2013. J Med Internet Res 2014;16(8):e195 [FREE Full text] [doi: 10.2196/jmir.3289] [Medline: 25158744]
20. Potts HW, Wyatt JC. Survey of doctors' experience of patients using the Internet. J Med Internet Res 2002;4(1):e5 [FREE Full text] [doi: 10.2196/jmir.4.1.e5] [Medline: 11956037]
21. Husson O, Denollet J, Oerlemans S, Mols F. Satisfaction with information provision in cancer patients and the moderating effect of Type D personality. Psychooncology 2013 Sep;22(9):2124-2132. [doi: 10.1002/pon.3267] [Medline: 23423829]
22. Eysenbach G. The impact of the Internet on cancer outcomes. CA Cancer J Clin 2003;53(6):356-371 [FREE Full text] [Medline: 15224975]
23. Lee CJ, Gray SW, Lewis N. Internet use leads cancer patients to be active health care consumers. Patient Educ Couns 2010;81:S63-S69.
24. van de Poll-Franse LV, Nicolaije KA, Vos MC, Pijnenborg JM, Boll D, Husson O, et al. The impact of a cancer Survivorship Care Plan on gynecological cancer patient and health care provider reported outcomes (ROGY Care): study protocol for a pragmatic cluster randomized controlled trial. Trials 2011;12:256 [FREE Full text] [doi: 10.1186/1745-6215-12-256] [Medline: 22141750]
25. Zwarenstein M, Treweek S, Gagnier JJ, Altman DG, Tunis S, Haynes B, Pragmatic Trials in Healthcare (Practihc) group. Improving the reporting of pragmatic trials: an extension of the CONSORT statement. BMJ 2008;337:a2390 [FREE Full text] [Medline: 19001484]
26. van de Poll-Franse LV, Horevoorts N, van Eenbergen M, Denollet J, Roukema JA, Aaronson NK, et al. The Patient Reported Outcomes Following Initial treatment and Long term Evaluation of Survivorship registry: scope, rationale and design of an infrastructure for the study of physical and psychosocial outcomes in cancer survivorship cohorts. Eur J Cancer 2011 Sep;47(14):2188-2194 [FREE Full text] [doi: 10.1016/j.ejca.2011.04.034] [Medline: 21621408]
27. Campbell M, Piaggio G, Elbourne D, Altman D. Consort 2010 statement: extension to cluster randomised trials. BMJ 2012 Sep 04;345(sep04 1):e5661-e5661. [doi: 10.1136/bmj.e5661]
28. Nicolaije KA, Ezendam NP, Vos MC, Pijnenborg JM, van de Poll-Franse LV, Kruitwagen RF. Oncology providers' evaluation of the use of an automatically generated cancer survivorship care plan: longitudinal results from the ROGY Care trial. J Cancer Surviv 2014 Jun;8(2):248-259. [doi: 10.1007/s11764-013-0327-1] [Medline: 24357226]
29. Ganz PA, Hahn EE. Implementing a survivorship care plan for patients with breast cancer. J Clin Oncol 2008 Feb 10;26(5):759-767. [doi: 10.1200/JCO.2007.14.2851] [Medline: 18258984]
30. Arraras JI, Greimel E, Sezer O, Chie W, Bergenmar M, Costantini A, et al. An international validation study of the EORTC QLQ-INFO25 questionnaire: an instrument to assess the information given to cancer patients. Eur J Cancer 2010 Oct;46(15):2726-2738. [doi: 10.1016/j.ejca.2010.06.118] [Medline: 20674333]
31. Brédart A, Bottomley A, Blazeby JM, Conroy T, Coens C, D'Haese S, European Organisation for Research and Treatment of Cancer Quality of Life GroupQuality of Life Unit. An international prospective study of the EORTC cancer in-patient satisfaction with care measure (EORTC IN-PATSAT32). Eur J Cancer 2005 Sep;41(14):2120-2131. [doi: 10.1016/j.ejca.2005.04.041] [Medline: 16182120]
32. Broadbent E, Petrie KJ, Main J, Weinman J. The brief illness perception questionnaire. J Psychosom Res 2006 Jun;60(6):631-637. [doi: 10.1016/j.jpsychores.2005.10.020] [Medline: 16731240]
33. Sangha O, Stucki G, Liang MH, Fossel AH, Katz JN. The Self-Administered Comorbidity Questionnaire: a new method to assess comorbidity for clinical and health services research. Arthritis Rheum 2003 Apr 15;49(2):156-163 [FREE Full text] [doi: 10.1002/art.10993] [Medline: 12687505]
34. Peters SA, Bots ML, den Ruijter HM, Palmer MK, Grobbee DE, Crouse JR, et al. Multiple imputation of missing repeated outcome measurements did not add to linear mixed-effects models. J Clin Epidemiol 2012 Jun;65(6):686-695. [doi: 10.1016/j.jclinepi.2011.11.012] [Medline: 22459429]
35. Beunckens C, Molenberghs G, Kenward MG. Direct likelihood analysis versus simple forms of imputation for missing data in randomized clinical trials. Clin Trials 2005;2(5):379-386. [Medline: 16315646]
36. Twisk J. Applied Multilevel Analysis: A Practical Guide for Medical Researchers. New York: Cambridge University Press; 2006.
37. Bickel R. Multilevel Analysis for Applied Research: It’s Just Regression!. New York: The Guilford Press; 2007.
38. Frazier P, Barron K, Tix A. Testing moderator and mediator effects in counseling psychology. J Couns Psychol 2004;51(1):115-134.
39. Adler NE, Newman K. Socioeconomic disparities in health: pathways and policies. Health Aff (Millwood) 2002;21(2):60-76 [FREE Full text] [Medline: 11900187]
40. Galobardes B, Shaw M, Lawlor DA, Lynch JW, Davey SG. Indicators of socioeconomic position (part 1). J Epidemiol Community Health 2006 Jan;60(1):7-12 [FREE Full text] [doi: 10.1136/jech.2004.023531] [Medline: 16361448]




XSL•FO
RenderX

41. Hill-Kayser CE, Vachani C, Hampshire MK, Jacobs LA, Metz JM. An internet tool for creation of cancer survivorship care plans for survivors and health care providers: design, implementation, use and user satisfaction. J Med Internet Res 2009;11(3):e39 [FREE Full text] [doi: 10.2196/jmir.1223] [Medline: 19734128]
42. Hill-Kayser CE, Vachani CC, Hampshire MK, Di Lullo G, Jacobs LA, Metz JM. Impact of internet-based cancer survivorship care plans on health care and lifestyle behaviors. Cancer 2013 Nov 1;119(21):3854-3860 [FREE Full text] [doi: 10.1002/cncr.28286] [Medline: 23922002]
43. Hill-Kayser CE, Vachani C, Hampshire MK, Metz JM. High level use and satisfaction with internet-based breast cancer survivorship care plans. Breast J 2012;18(1):97-99. [doi: 10.1111/j.1524-4741.2011.01195.x] [Medline: 22098063]
44. Statline Statistics Netherlands. 2015. Statline: ICT gebruik van huishoudens naar huishoudkenmerken URL: http://statline.cbs.nl/StatWeb/publication/?VW=T&DM=SLNL&PA=71102ned&D1=0-3,30-34&D2=0-10&D3=0,l&HD=130422-1118&HDR=G2,T&STB=G1 [WebCite Cache ID 6hhTLVZs8]
45. Husson O, Thong MS, Mols F, Oerlemans S, Kaptein AA, van de Poll-Franse LV. Illness perceptions in cancer survivors: what is the role of information provision? Psychooncology 2013 Mar;22(3):490-498. [doi: 10.1002/pon.3042] [Medline: 22307579]
46. Ezendam NP, Thong MS, Nicolaije KA, Kruitwagen RF. ISurvivorship Care Plans to inform the primary care physician: results from the ROGY care pragmatic cluster randomized controlled trial.. J Cancer Surviv 2014(8):595-602. [doi: 10.1007/s11764-014-0368-0] [Medline: 24866923 ]

## Abbreviations

**B-IPQ:** Brief Illness Perception Questionnaire
**EORTC:** European Organisation for Research and Treatment of Cancer
**FIGO:** Federation of Gynecology and Obstetrics
**IOM:** Institute of Medicine
**ROGY:** Registrationsystem Oncological GYnecology
**SCP:** Survivorship Care Plan
**SES:** socioeconomic status

*Edited by G Eysenbach; submitted 08.07.15; peer-reviewed by S Birken; comments to author 14.10.15; revised version received 26.11.15; accepted 19.01.16; published 08.07.16.*

*Please cite as:*
*Nicolaije KAH, Ezendam NPM, Pijnenborg JMA, Boll D, Vos MC, Kruitwagen RFPM, van de Poll-Franse LV*
*Paper-Based Survivorship Care Plans May be Less Helpful for Cancer Patients Who Search for Disease-Related Information on the Internet: Results of the Registrationsystem Oncological Gynecology (ROGY) Care Randomized Trial*
*J Med Internet Res 2016;18(7):e162*
*URL: http://www.jmir.org/2016/7/e162/*
*doi:10.2196/jmir.4914*
*PMID:27392550*

©Kim AH Nicolaije, Nicole PM Ezendam, Johanna MA Pijnenborg, Dorry Boll, Maria Caroline Vos, Roy FPM Kruitwagen, Lonneke V van de Poll-Franse. Originally published in the Journal of Medical Internet Research (http://www.jmir.org), 08.07.2016. This is an open-access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/2.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on http://www.jmir.org/, as well as this copyright and license information must be included.



XSL•FO
RenderX

Original Paper

# Sustained Use of Patient Portal Features and Improvements in Diabetes Physiological Measures

Stephanie L Shimada[1,2,3], PhD; Jeroan J Allison[3], MD, MS; Amy K Rosen[4,5], PhD; Hua Feng[1,3], MS, PhD; Thomas K Houston[1,3], MD, MPH

[1]Center for Healthcare Organization and Implementation Research (CHOIR), Edith Nourse Rogers Memorial Veterans Hospital, Bedford, MA, United States

[2]Department of Health Law, Policy, and Management, Boston University School of Public Health, Boston, MA, United States

[3]Division of Health Informatics and Implementation Science, Department of Quantitative Health Sciences, University of Massachusetts Medical School, Worcester, MA, United States

[4]Center for Healthcare Organization and Implementation Research (CHOIR), VA Boston Healthcare System, Boston, MA, United States

[5]Department of Surgery, Boston University School of Medicine, Boston, MA, United States

**Corresponding Author:**
Stephanie L Shimada, PhD
Center for Healthcare Organization and Implementation Research (CHOIR)
Edith Nourse Rogers Memorial Veterans Hospital
200 Springs Rd, Building 70 (152)
Bedford, MA, 01730
United States
Phone: 1 781 687 2208
Fax: 1 781 687 3106
Email: Stephanie.Shimada@va.gov

## Abstract

**Background:** Personal health records (PHRs) have the potential to improve patient self-management for chronic conditions such as diabetes. However, evidence is mixed as to whether there is an association between PHR use and improved health outcomes.

**Objective:** The aim of this study was to evaluate the association between sustained use of specific patient portal features (Web-based prescription refill and secure messaging—SM) and physiological measures important for the management of type 2 diabetes.

**Methods:** Using a retrospective cohort design, including Veterans with diabetes registered for the My Health *e* Vet patient portal who had not yet used the Web-based refill or SM features and who had at least one physiological measure (HbA1c, low-density lipoprotein (LDL) cholesterol, blood pressure) in 2009-2010 (baseline) that was above guideline recommendations (N=111,686), we assessed portal use between 2010 and 2014. We calculated the odds of achieving control of each measure by 2013 to 2014 (follow-up) by years of using each portal feature, adjusting for demographic and clinical characteristics associated with portal use.

**Results:** By 2013 to 2014, 34.13% (38,113/111,686) of the cohort was using Web-based refills, and 15.75% (17,592/111,686) of the cohort was using SM. Users were slightly younger ($P$<.001), less likely to be eligible for free care based on economic means ($P$<.001), and more likely to be women ($P$<.001). In models adjusting for both features, patients with uncontrolled HbA1c at baseline who used SM were significantly more likely than nonusers to achieve glycemic control by follow-up if they used SM for 2 years (odds ratio—OR=1.24, CI: 1.14-1.34) or 3 or more years (OR=1.28, CI: 1.12-1.45). However, there was no significant association between Web-based refill use and glycemic control. Those with uncontrolled blood pressure at baseline who used Web-based refills were significantly more likely than nonusers to achieve control at follow-up with 2 (OR=1.07, CI: 1.01-1.13) or 3 (OR=1.08, CI: 1.02-1.14) more years of Web-based refill use. Both features were significantly associated with improvements in LDL cholesterol levels at follow-up.

**Conclusions:** Although rates of use of the refill function were higher within the population, sustained SM use had a greater impact on HbA1c. Evaluations of patient portals should consider that individual components may have differential effects on health improvements.



XSL•FO
RenderX

*(J Med Internet Res 2016;18(7):e179)* doi:10.2196/jmir.5663

**KEYWORDS**

personal health records; diabetes mellitus, type 2; self care; HbA1c; cholesterol, LDL; blood pressure; veterans

## Introduction

Diabetes affects over 29 million Americans [1] and was estimated to have cost between $245 billion [2] and $322 billion [3] in 2012. Despite advances in effective treatments [4], almost half of those with type 2 diabetes do not meet recommended targets for glycemic control, low-density lipoprotein (LDL) cholesterol control, or blood pressure control [5]. Poor control of diabetes is associated with poor health outcomes, increased morbidity, and mortality [1,3]. Type 2 diabetes affects a large portion of US Veterans, with 25% of Veterans having the diagnosis [6,7].

Patients with diabetes and other chronic diseases do not do well with episodic, transactional care limited to in-person visits. The Institute of Medicine [8] has called for a shift toward continuous, coordinated care, leveraging information technology to support self-management and communication between clinic visits. Type 2 diabetes requires patient self-management and effective patient–provider communication to tailor treatments, manage side effects, monitor physiological processes, and screen for complications. Personal health records (PHRs) and patient portals are technologies with the potential to increase patient self-management and enable patients to better communicate with their clinical teams [9,10].

Evidence for patient portal effectiveness for chronic disease management is limited, and association with outcomes is mixed [11]. Tenforde et al [12] found that portal use was associated with improvements in diabetes-related quality measures but did not find a dose-response association with varying intensity of portal use and did not separate out effects by specific portal feature. Potential benefits of portal use have included patient reports of enhanced satisfaction, improved access outside of face-to-face visits, and improved efficiency and quality of face-to-face visits [13]. Studies from Kaiser [14] and Group Health Cooperative [15,16] found significant associations between use of secure messaging (SM) and improvements in diabetes care, with significant performance improvements in glycemic testing and control. Other studies have documented improvements in medication adherence among diabetic patients on statins exclusively using Web-based prescription refill through a patient portal [17]. Association of portal use and improvements in cholesterol and blood pressure effectiveness of care measures [14] have also been documented among patients with diabetes and hypertension.

Portals vary widely, adding to the difficulty in evaluating any effects they may have on patients' health outcomes. Some are tethered to a health care system, others are not, some are disease specific, whereas most are not [18,19]. The Department of Veterans' Affairs provides its patients with a portal, My HealtheVet (MHV), including features allowing them to refill VA prescriptions and send secure messages to their providers [20-22]. These two features, SM and Web-based prescription refill, are among the most common across portals and are the most frequently used [23]. Veterans with diabetes have relatively high adoption of MHV and of these key features [7].

This study examines whether diabetes outcomes are improved for patients with type 2 diabetes who initiate use of key features of the MHV patient portal compared with similar patients with type 2 diabetes who are also registered for the portal but do not initiate use of any of these features. To answer this question, we focused on patients with a diagnosis of type 2 diabetes who had at least one uncontrolled physiological measure (hemoglobin A1c, LDL cholesterol, blood pressure) at baseline (2009-2010) to examine whether those who had used the portal's Web-based prescription refill or SM features for the first time between 2010 and 2013 were more likely than nonusers to achieve control at follow-up (2013-2014). We also sought to explore both the separate and combined effects of Web-based refill and SM use on physiological measures and whether sustained use was associated with a greater probability of achieving control.

## Methods

### Study Design and Overview

We conducted a 5-year retrospective cohort study of Veterans with type 2 diabetes registered for the MHV portal. Data for these analyses came from the Veteran's Health Administration's Corporate Data Warehouse, including administrative data, clinical records for inpatient and outpatient care, and MHV registration and use data. We used International Classification of Diseases, Ninth Revision, Clinical Modification (ICD-9-CM) diagnosis codes (October 1, 2007-March 31, 2009) to determine type 2 diabetes diagnosis and determine patient characteristics at baseline. Data from April 1, 2009 through March 31, 2014 were used to assess MHV use over time. Intermediate physiological measures obtained during clinical care were obtained at baseline and follow-up. In addition, we linked income and educational attainment variables from the US Census Bureau's 2007- 2011 American Community Survey (5-year estimates) to each Veteran via postal code.

### Cohort Eligibility

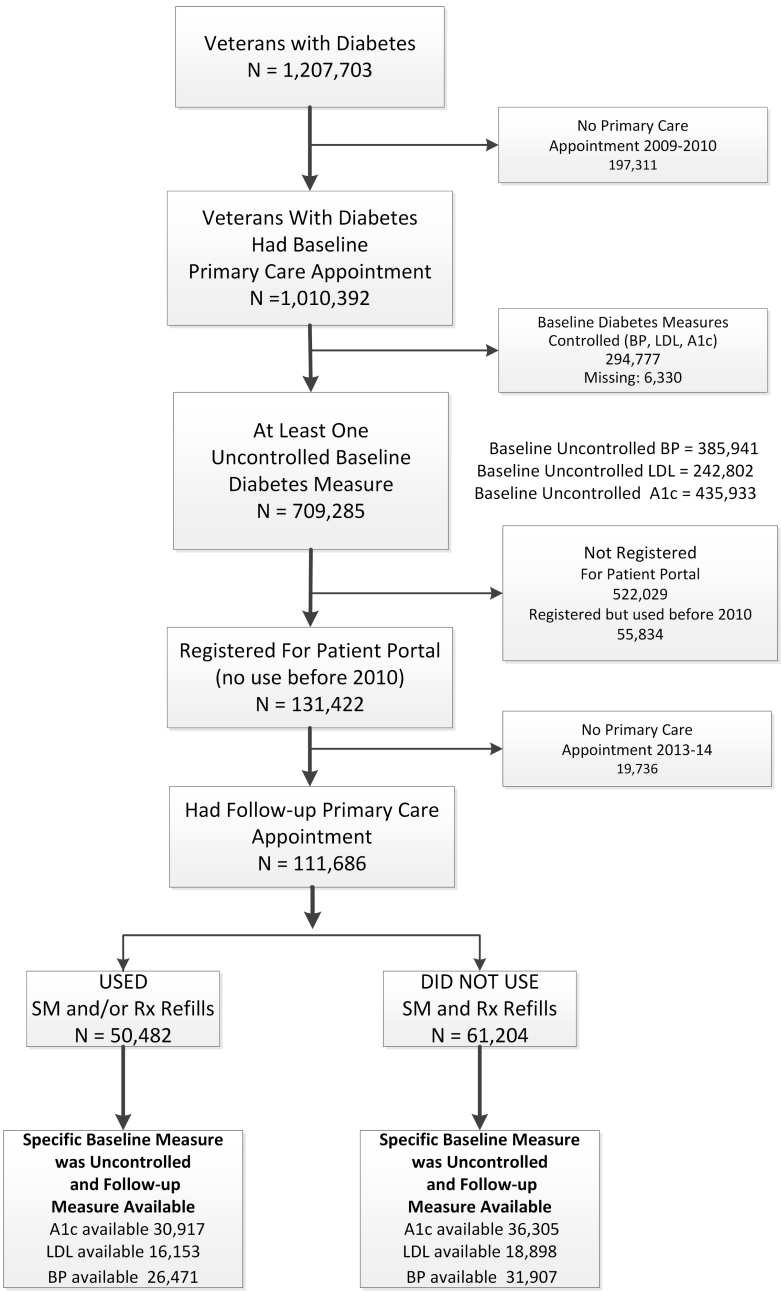
We identified patients who had at least two outpatient records or one inpatient record with an ICD-9-CM diagnosis code for type 2 diabetes by March 2009 (N=1,207,703). Use of two or more diabetes-related ICD-9-CM codes from inpatient or outpatient visits has previously been determined to be the most accurate way to identify patients with diabetes in VA administrative data [24]. We then excluded patients who had not used the VA for primary care in 2009 to 2010, who had controlled or missing diabetes outcome measures, who were not registered for the portal, or who had used the MHV Web-based prescription refill or SM features before 2010 (see Figure 1). We limited our analyses to those who were registered to use the MHV portal to minimize differences in access to the portal or in willingness to use the portal among users and



XSL•FO
RenderX

nonusers so that we could focus on associations with actual use. Our previous work has shown patients registered for the portal (but not using features) to be a more appropriate and comparable reference group [7]. Because our goal was to understand how a patient portal could assist in achieving improvements in physiological control, we also excluded those who were controlled at baseline from the main analyses as those patients had already successfully managed to control their physiological measures without the use of MHV. The final analysis cohort included 111,686 patients.

**Figure 1.** Cohort Selection.




## Variables

### *Dependent Variables—Diabetes-Related Physiological Measures: HbA1c, LDL, Systolic and Diastolic Blood Pressures*

We used the American Diabetes Association's guidelines to define cutoffs for glycemic, cholesterol, and blood pressure control [25]. We defined uncontrolled blood glucose at baseline (April 2009-March 2010) if the patient's average hemoglobin A1c (HbA1c) during that period was greater than or equal to 7.0% (53 mmol/mol) and uncontrolled LDL cholesterol at baseline if the patient's average cholesterol reading during that same period was greater than or equal to 100 mg/dL (2.586 mmol/L). Similarly, we determined that a patient had uncontrolled blood pressure if the average systolic blood pressure at baseline (assessed by averaging all readings during the baseline year) was 140 mmHg or higher, or the average diastolic blood pressure (similarly averaged across the baseline year) was 80 mmHg or higher. To achieve control by follow-up, patients had to lower their readings to below the cutoffs (blood glucose and LDL cholesterol) or achieve control over both systolic and diastolic blood pressures (blood pressure). A binary



XSL•FO
RenderX

indicator for whether a patient with uncontrolled values at baseline achieved control by follow-up (2013-2014) was the dependent variable for the logistic regression models.

### *Independent Variables—Use of the Portal: Registration, Use of Web-Based Prescription Refill and Secure Messaging*

Among Veterans registered by April 1, 2013, we measured use of two key features of the portal, which had been available throughout the study period: the Web-based prescription refill feature and the SM feature and used a binary indicator of any use to describe the samples. We assessed how often each patient used each feature during the potential exposure period (April 2010-March 2013). At some facilities, patients were prompted to try these features (eg, send a test message to one's primary care team) as part of a MHV training. We therefore defined "use" as two or more prescriptions filled online via the MHV portal per year or two or more SMs sent per year, to ensure we captured actual use and not just attendance at a training session. To measure dose of exposure, our primary measure of use for each portal feature was a categorical variable indicating whether a patient had used each feature two or more times per year over 1 year, 2 years, or for 3 or more years during the potential exposure period. A continuous variable measuring years of use (ie, years with 2+ refills or 2+ SMs sent) for each portal feature was used for tests for trend.

### *Other Covariates*

Other covariates we used included demographic characteristics such as patient age, gender, race or ethnicity, urban, suburban, or rural residence, educational attainment, and income. In multivariable models, we adjusted for age, gender, race, comorbidities, and available measures of socioeconomic status because these have been significantly associated with adoption of SM and patient portals in previous studies [9,26]. For income, we included a measure of whether the patient was eligible for free care from the Veterans Health Administration based on low income. Because data on Veterans' income and educational attainment do not exist in the VA Corporate Data Warehouse, we also linked Census data by postal code of residence on the percentage of adults aged older than 25 years who have attained a bachelor's degree or higher and the median per capita earnings in the past 12 months (in 2011 inflation-adjusted dollars) among those aged 25 years and older with earnings. We also adjusted for the number of primary care visits a patient had during the baseline year and the number of comorbidities they had as determined by the Elixhauser algorithm for identifying comorbidities from administrative data [27].

### Analyses

We characterized the overall cohort and examined means and distributions of patient demographic and clinical characteristics by use, both overall and for those with specific uncontrolled physiological measures at baseline. We calculated the proportion of patients with diabetes in our cohort using each feature over each year of the study and the average number of prescriptions refilled or secure messages sent during each year. Our primary goal was to assess the association of use of patient portal features with change in diabetes-relevant physiological measures (HbA1c, LDL, BP). To achieve this goal, we first calculated means and binomial confidence intervals (CIs) for the proportion of patients who were uncontrolled at baseline who achieved control at follow-up, stratified by the number of years of use of the SM or Web-based refill features. We then constructed a series of logistic regression models predicting control of each physiological measure at follow-up based on categorical measures of portal use (years of use of each feature), adjusting for the covariates described previously. All logistic regression models were adjusted for patient age; gender; race or ethnicity; eligibility for free VA health care; number of Elixhauser comorbidities at baseline; number of primary care visits at baseline (in 2009-2010) and during the study period (2010-2014); urban, suburban, rural, or highly rural residence; median income by postal code; and the percentage of college graduates in the patient's residential postal code. In addition, models for control of blood pressure, cholesterol, and HbA1c at follow-up (2013-2014) were adjusted for the patient's mean baseline blood pressure, LDL cholesterol, or HbA1c value in 2009 to 2010, respectively. Separate models were first run for each feature (Web-based prescription refill use and SM use) because there was a moderate correlation between uses of the two features. To further evaluate the independent effect of each feature, we also ran combined logistic models, which included both Web-based prescription refill use and SM use in the same models. To test for dose response, we then ran tests of trend treating the number of years of use of each feature as a continuous variable. We also conducted sensitivity analyses to see whether results changed depending on (1) our definition of use (ie, defining use as one or more uses of a feature in a given year) or (2) inclusion of patients who met other inclusion criteria but were controlled at baseline in the analysis sample.

## Results

### Feature Use

Within our cohort of 111,686 patients (see Figure 1), 50,482 (45.20%) used Web-based prescription refills or SM or both at least twice per year between April 2010 and March 2014, and 61,204 (54.80%) used neither.

### Patient Characteristics

Table 1 describes the characteristics of the overall sample and examines differences between patients who used the Web-based prescription refill feature or the SM feature or both in MHV between April 2010 and March 2014 and those who did not.



XSL•FO
RenderX

**Table 1.** Characteristics of patients with type 2 diabetes registered for My HealtheVet, overall and by use or nonuse of the Web-based refill or secure messaging features as of March 2014.

| Variables | Overall | Used neither Web-based refill nor SM as of March 2014 (nonusers) | Used Web-based refill or SM or both as of March 2014 (users) | Difference between user and nonuser groups (Pearson's chi-square or 2-sided *t*-test) |
|---|---|---|---|---|
| **N** | 111,686 | 61,204 | 50,482 | |
| **Age (mean (SD)** | 62.05 (9.6) | 63.22 (9.6) | 60.63 (9.5) | $t_{111684}$ = 45.2, *P*<.001 |
| **Gender (% female)** | 3.58% | 3.16% | 4.08% | $\chi^2_1$ = 67.2, *P*<.001 |
| **Race or ethnicity** | | | | |
| White | 68.87% | 67.25% | 70.84% | Reference group for $\chi^2$ |
| African-American | 16.95% | 18.64% | 14.90% | $\chi^2_1$ = 280.3, *P*<.001 |
| Latino | 5.69% | 5.63% | 5.76% | $\chi^2_1$= 1.1, *P*=.294 |
| Native Hawaiian or Pacific Islander | 1.10% | 1.08% | 1.12% | $\chi^2_1$ = 0.1, *P*=.755 |
| Asian | 0.81% | 0.73% | 0.90% | $\chi^2_1$ = 5.9, *P*=.015 |
| American Indian or Alaska Native | 0.71% | 0.73% | 0.70% | $\chi^2_1$ = 1.8, *P*=.178 |
| Unknown to patient, refused, or missing | 5.87% | 5.94% | 5.79% | $\chi^2_1$ = 9.1, *P*=.003 |
| **Percent eligible for free VA health care based on income** | 20.43% | 21.71% | 18.87% | $\chi^2_1$ = 137.6, *P*<.001 |
| **Median income in postal code of residence US$ (mean (SD))** | 33,548.86 (8,926.98) | 33,532.67 (8,996.24) | 33,568.51 (8,842.24) | $t_{108985}$ = −0.6596, *P*=.5095 |
| **Percent of adults with a university degree or higher in postal code of residence (mean (SD))** | 23.46% (12.7 ) | 23.39% (12.8) | 23.54% (12.7) | $t_{109086}$= −1.9398, *P*=.0524 |
| **Location** | | | | |
| Urban (%) | 73.23% | 73.27% | 73.19% | Reference group for $\chi^2$ |
| Suburban (%) | 13.40% | 13.46% | 13.32% | $\chi^2_1$ = 0.3, *P*=.597 |
| Rural (%) | 7.23% | 7.32% | 7.12% | $\chi^2_1$ = 1.3, *P*=.259 |
| Highly rural (%) | 6.14% | 5.95% | 6.38% | $\chi^2_1$ = 7.5, *P*=.006 |
| **Number of Elixhauser comorbidities at baseline (mean (SD))** | 5.57 (2.5) | 5.56 (2.6) | 5.59 (2.5) | $t_{111357}$=−1.826, *P*=.0679 |
| **Number of primary care visits at baseline (mean(SD))** | 4.40 (3.6) | 4.37 (3.6) | 4.43 (3.6) | $t_{111684}$ = −3.0046, *P*=.0027 |
| **Number of primary care visits from 2010 to 2014 (mean (SD))** | 17.59 (12.8) | 16.97 (12.6) | 18.34 (13.0) | $t_{111684}$ = −17.86, *P*<.001 |

Compared with patients who did not use either of the features, patients who used Web-based refill or SM were slightly younger (60.6 years vs 63.2 years, *P*<.001), more likely to be female (4.08% vs 3.16%, *P*<.001), and less likely to be eligible for free VA care based on low economic means (18.87% vs 21.71%, *P*<.001). There were significant differences in race or ethnicity between users and nonusers, with African-American (*P*<.001), Asian (*P*=.015), and patients of unknown race (*P*=.003) less likely than white patients to be users. The difference was most marked between African-American and white patients (39.73% (7,521/18,931) of African-American patients were users vs 46.49% (35,759/76,920) of white patients, *P*<.001). Although most patients resided in urban areas, urban patients were slightly less likely to be users than patients residing in areas designated as highly rural (45.06% (36,078/80,060) vs 46.80% (3,143/6,716), *P*=.006).

There were no significant differences in the number of Elixhauser comorbidities at baseline (*P*=.0679), median income



XSL•FO
RenderX

by postal code of residence (*P*=.5095), percentage of adults with a bachelor's degree or higher in postal code of residence (*P*=.0524). There was a statistically significant difference in the number of primary care visits at baseline in the overall cohort (4.37 visits for nonusers vs 4.43 visits for users at baseline, *P*=.0027), but this difference vanished when looking at analysis subgroups based on uncontrolled measure at baseline (see Table 2). There was a highly significant difference in the number of primary care visits between 2010 and 2014 (16.97 for nonusers vs 18.34 visits for users, *P*<.001). Users also showed evidence of higher primary care utilization in all analysis subgroups (see Table 2).

Further detail describing the characteristics based on each uncontrolled measure (ie, the sample for each logistic regression model) is summarized in Table 2.

**Table 2.** Demographics of patients with type 2 diabetes registered for My HealtheVet by uncontrolled physiological measure at baseline and by use of the portal.

| Variables | Uncontrolled Measure | | | | | |
|---|---|---|---|---|---|---|
| | Hemoglobin A1c A1c ≥7.0% | | Low-density Lipoprotein LDL≥100mg/dL | | Blood Pressure BP≥140/80 mmHg | |
| | Registered, no use | Used SM or Web-based refill | Registered, no use | Used SM or Web-based refill | Registered, no use | Used SM or Web-based refill |
| **N** | 36,305 | 30,917 | 18,898 | 16,153 | 31,907 | 26,471 |
| **Age (mean (SD)** | 62.66 (9.2) | 60.28 (9.2) | 61.47 (9.6) | 58.68 (9.6) | 62.63 (9.7) | 60.13 (9.7) |
| **Gender (% female)** | 2.86% | 3.71% | 4.87% | 6.53% | 2.88% | 3.58% |
| **Race or ethnicity** | | | | | | |
| White | 66.88% | 70.98% | 64.21% | 67.62% | 64.55% | 68.44% |
| African-American | 19.12% | 14.84% | 22.14% | 17.56% | 21.38% | 17.25% |
| Latino | 6.08% | 6.03% | 5.79% | 6.27% | 5.53% | 5.63% |
| Native Hawaiian Pacific Islander | 1.08% | 1.11% | 1.04% | 1.08% | 1.08% | 1.21% |
| Asian | 0.75% | 0.89% | 0.71% | 1.00% | 0.72% | 0.87% |
| American Indian or Alaska Native | 0.73% | 0.70% | 0.69% | 0.84% | 0.71% | 0.70% |
| Unknown to patient, refused, or missing | 5.36% | 5.45% | 5.42% | 5.64% | 6.02% | 5.91% |
| **Percent eligible for free VA health care** | 22.38% | 19.40% | 21.61% | 18.76% | 21.93% | 18.96% |
| **Median income in postal code US$ (mean (SD))** | 33,453.54 (8925.47) | 33,548.58 (8839.33) | 33,111.41 (8813.43) | 33,197.85 (8674.04) | 33,364.98 (8815.67) | 33,424.07 (8763.60) |
| **Percent adults with a university degree or higher in postal code (mean(SD))** | 23.12% (12.7) | 23.32% (12.6) | 22.95% (12.6) | 23.20% (12.4) | 23.27% (12.7) | 23.46% (12.6) |
| **Location** | | | | | | |
| Urban (%) | 73.21% | 73.12% | 73.32% | 73.43% | 73.40% | 72.91% |
| Suburban (%) | 13.34% | 13.15% | 13.64% | 13.15% | 13.54% | 13.47% |
| Rural (%) | 7.41% | 7.24% | 7.18% | 7.05% | 7.14% | 7.12% |
| Highly rural (%) | 6.04% | 6.50% | 5.86% | 6.37% | 5.93% | 6.50% |
| **Number Elixhauser comorbidities at baseline (mean (SD))** | 5.70 (2.6) | 5.72 (2.5) | 5.40 (2.5) | 5.45 (2.4) | 5.54 (2.5) | 5.51 (2.4) |
| **Number of primary care visits at baseline (2009-10; mean(SD))** | 4.64 (3.8) | 4.66 (3.7) | 4.32 (3.5) | 4.38 (3.5) | 4.29 (3.5) | 4.30 (3.4) |
| **Number of primary care visits; 2010 to 2014, mean (SD)** | 18.04 (12.9) | 19.17 (13.3) | 17.09 (12.4) | 18.29 (12.7) | 16.97 (12.5) | 18.17 (12.5) |

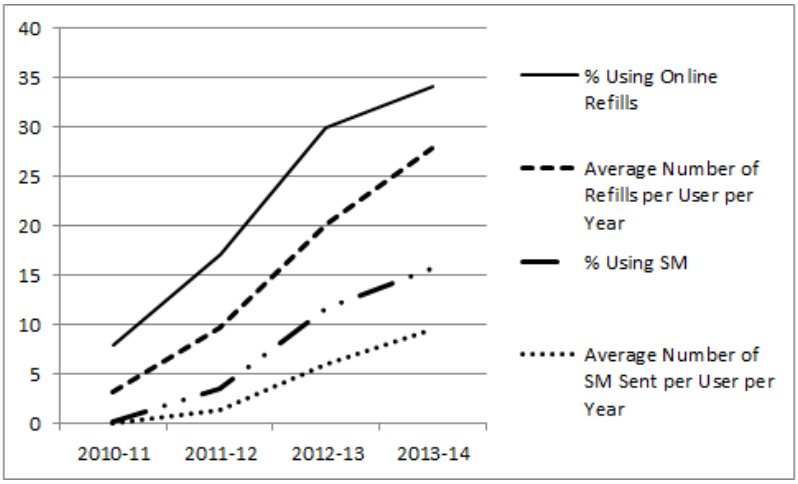


XSL•FO
RenderX

### Portal Use

Use of Web-based refills and SM increased steadily from 2010 to 2014 (Figure 2). Among registered patients with diabetes who had not used the portal before 2010, only 7.98% (8,917/111,686) used Web-based prescription refills in 2010 to 2011, and the average number of refills per year was 3.13 per user. In the same year, as SM was just being implemented at most facilities, only 0.22% (241/111,686) used SM and sent an average of 0.059 messages per user. By 2013 to 2014, the numbers had risen to 34.13% (38,113/111,686) of new users using Web-based refills, filling an average of 27.84 prescriptions each, and 15.75% (17,592/111,686) were using SM, sending an average of 9.46 messages each.

**Figure 2.** Proportion of patients with type 2 diabetes registered for My HealtheVet and first using Web-based prescription refills or secure messaging after 2010, increase in feature adoption over time, and average number of uses per user per year.




### Association of Patient Portal Use and Change in Diabetes Physiological Measures over 5 Years

The logistic regression results are presented in Table 3.

Our single-feature logistic regression models (Models 1a-c and Models 2a-c) showed that patients with uncontrolled HbA1c at baseline (2009-2010) were significantly more likely to achieve glycemic control at follow-up (2013-2014) if they used SM for 2 or more years. The odds of having an HbA1c below 7.0% (53 mmol/mol) at follow-up were 22% higher (after 2 years of use, odds ratio: OR=1.22, CI: 1.13-1.32) and 28% higher (after 3 or more years, OR=1.28, CI: 1.13-1.44), for those using SM compared with those who never used it.

However, use of Web-based prescription refills was only associated with glycemic control at follow-up after 3 or more years of use (OR=1.07, CI: 1.01-1.14). Those with uncontrolled blood pressure at baseline were significantly more likely to achieve control at follow-up only with 2 (OR=1.06, CI: 1.01-1.12) or 3 or more (OR=1.05, CI: 1.00-1.11) years of Web-based refill use, compared with nonusers. Use of SM was not significantly associated with improvements in blood pressure control. Both Web-based refill use and SM use were significantly associated with improvements in LDL cholesterol levels at follow-up. Compared with nonusers, the odds of users having LDL cholesterol below 100 mg/dL (2.586 mmol/L) were 12% higher with 2 years of Web-based refill use (OR=1.12, CI: 1.05-1.20), 16% higher with 3+ years of Web-based refill Use (OR=1.16, CI: 1.08-1.24), 9% higher with 1 year of SM use (OR=1.09, CI: 1.01-1.18), 17% higher with 2 years of SM use (OR=1.17, CI: 1.07-1.27), and 22% higher with 3+ years of SM use (OR=1.22, CI: 1.06-1.40).

We also ran logistic regression models identical to those mentioned previously that included both years of SM and Web-based refill use in the same model (Models 3a-c), as well as logistic regression models that included years of SM or Web-based refill use as a continuous variable as a test for trend (Models 4a-c). The conclusions remained largely unchanged, although ORs for the association between SM use and LDL were more attenuated (and no longer significant with the exception of 2 years of SM use) in the combined model. The combined model (and test for trend) did not show a significant association between SM use and blood pressure control ($P$=.370 for trend), or between Web-based refill use and glycemic control ($P$=.585 for trend); however, tests for trend revealed significant dose-response relationships between use of SM and glycemic control ($P$<.001), use of Web-based refill and blood pressure control ($P$=.001), and use of both features and LDL control ($P$<.001 and $P$=.015 for trend, respectively, for refills and SM use).



XSL•FO
RenderX

**Table 3.** Adjusted odds of being in control at follow-up (OR (95% CI)) for a patient with uncontrolled physiological measures (HbA1c, LDL, or blood pressure) at baseline, based on years of portal feature use.

| Models[a] | Health Outcomes in 2013-14 | | |
|---|---|---|---|
| | **Hemoglobin A1c** | **Low-density Lipoprotein** | **Blood Pressure** |
| | HbA1c<7% (53 mmol/mol) | LDL < 100 mg/dL (2.586 mmol/l) | SBP<140 mmHg DBP<80 mmHg |
| **Models 1a-c: adjusted odds ratios (95% CI) for being controlled in 2013-2014 among patients with uncontrolled physiological measures in 2009-10 based on years of Web-based prescription refill use[a]** | | | |
| Web-based prescription refill use | | | |
| None | Reference | Reference | Reference |
| 1 year | 0.99 (0.93, 1.05) | 1.01 (0.95, 1.08) | 1.02 (0.97, 1.08) |
| 2 years | 1.01 (0.95, 1.08) | 1.12 (1.05, 1.20)[c] | 1.06 (1.01, 1.12)[b] |
| 3 or more years | 1.07 (1.01, 1.14)[b] | 1.16 (1.08, 1.24)[d] | 1.05 (1.00, 1.11)[b] |
| **Models 2a-c: adjusted odds ratios (95% CI) for being controlled in 2013-2014 among patients with uncontrolled physiological measures in 2009-2010 based on years of secure messaging use[a]** | | | |
| Secure messaging use | | | |
| None | Reference | Reference | Reference |
| 1 year | 1.03 (0.96, 1.10) | 1.09 (1.01, 1.18)[b] | 1.03 (0.97, 1.09) |
| 2 years | 1.22 (1.13, 1.32)[d] | 1.17 (1.07, 1.27)[c] | 1.03 (0.96, 1.10) |
| 3 or more years | 1.28 (1.13, 1.44)[d] | 1.22 (1.06, 1.40)[c] | 1.00 (0.90, 1.12) |
| **Models 3a-c: adjusted odds ratios (95% CI) for being controlled in 2013-2014 among patients with uncontrolled physiological measures in 2009-2010 based on years of both features[a]** | | | |
| Web-based prescription refill use | | | |
| None | Reference | Reference | Reference |
| 1 year | 0.96 (0.91, 1.03) | 1.01 (0.94, 1.08) | 1.02 (0.97, 1.07) |
| 2 years | 0.96 (0.90, 1.03) | 1.13 (1.05, 1.21)[c] | 1.07 (1.01, 1.13)[b] |
| 3 or more years | 1.00 (0.94, 1.07) | 1.13 (1.05, 1.22)[c] | 1.08 (1.02, 1.14)[c] |
| Secure messaging use | | | |
| None | Reference | Reference | Reference |
| 1 year | 1.04 (0.97, 1.12) | 1.05 (0.97, 1.14) | 1.00 (0.94, 1.07) |
| 2 years | 1.24 (1.14, 1.34)[d] | 1.10 (1.00, 1.21)[b] | 0.98 (0.91, 1.05) |
| 3 or more years | 1.28 (1.12, 1.45)[d] | 1.12 (0.96, 1.30) | 0.95 (0.85, 1.07) |
| **Models 4a-c: combined tests for trend predicting controlled outcomes in 2013-2014 among patients with uncontrolled physiological measures in 2009-2010 based on years of use for both features[a]** | | | |
| Web-based prescription refill use | *P*=.585 | *P*<.001 | *P*=.001 |
| Secure messaging use | *P*<.001 | *P*=.015 | *P*=.370 |

[a]All models adjust for patient characteristics in Table 1 including age, gender, race or ethnicity, eligibility for free care, geographic location, number of Elixhauser comorbidities, and baseline number of primary care visits in 2009 to 2010. In addition, models adjusted for the patient's physiological measure (blood pressure, LDL cholesterol, or HbA1c value) in 2009 to 2010, median income in the patient's residential zip code, and the percentage of college graduates in the patient's residential postal code.

[b]Odds ratios are significant at the *P*<.05 level as indicated.

[c]Odds ratios are significant at the *P*<.01 level as indicated.

[d]Odds ratios are significant at the *P*<.001 level as indicated.



XSL•FO
RenderX

**Figure 3.** Proportion controlled at follow-up, out of all diabetics uncontrolled for that specific measure at baseline (proportion and binomial CIs).

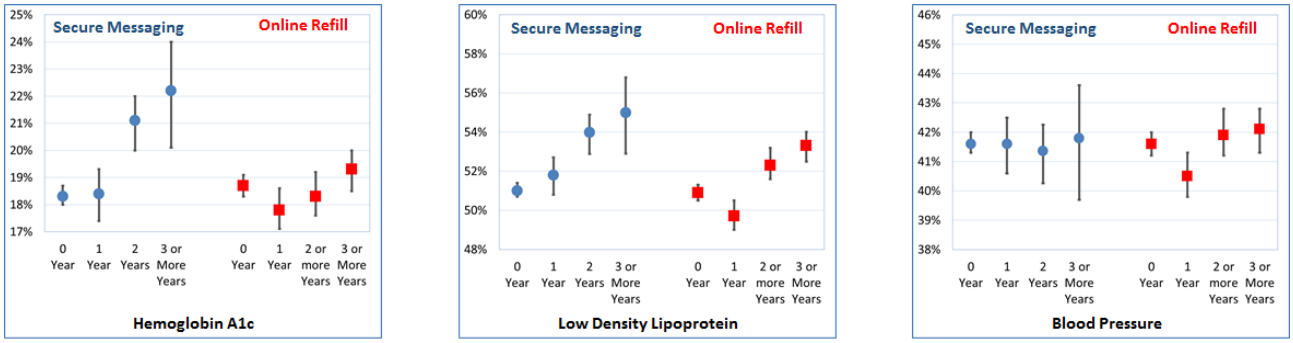



Figure 3 shows the predicted probability of achieving control over each measure by follow-up based on years of refill and SM use among those uncontrolled at baseline for each measure. The figure illustrates how sustained use of each tool is associated with improvements in control of physiological measures.

### Sensitivity Analyses

We conducted sensitivity analyses to see whether our results would change with the inclusion of those whose physiological measures were controlled at baseline, but otherwise met criteria for inclusion. Although the ORs were attenuated, significant tests for trend revealed the same relationships between feature use and being in control at follow-up for all the measures. Similarly, when use was defined as use of a feature even once in a given year, ORs were again somewhat attenuated; however, the results, including the tests for trend, led to identical conclusions about the associations between feature use and controlled physiological outcomes at follow-up.

## Discussion

### Principal Findings

Within this cohort of patients with type 2 diabetes and uncontrolled physiological measures, we saw increasing activity on the MHV patient portal between 2010 and 2014. The rate of use and increase in use was greater for Web-based refills than for SM. We observed small, statistically significant, and potentially meaningful improvement in physiological measures among diabetic patients who initiated and sustained use of Web-based refills or SM or both via MHV. However, the association varied by specific MHV feature. Where a significant association was found, use of SM was associated with higher odds of improved outcomes than use of Web-based refills.

### Comparison With Prior Work

The association between use of SM and improved diabetes physiological measures is consistent with that of prior research [14-16]; however, we were able to add information on the effects of sustained use over many years. For most measures, we found a dose-response effect on outcomes, suggesting that sustained use of the feature was associated with greater likelihood of being controlled at follow-up. The more years the patient used the feature, the greater the odds of achieving control compared with those who did not use the feature. Use of SM was associated with improvements in glycemic control with sustained use over 2 to 3+ years. Type 2 diabetic patients with uncontrolled blood pressure were more likely to achieve blood pressure control with 2 to 3+ years' use of Web-based medication refills through MHV. Both prescription refills and SM were associated with improvements in lipid levels with sustained use. Adjusting for use of both the features in the model did shift the magnitude of the odds of achieving control. This suggests that the association between patient portal use and health outcomes will vary based on the combination of different features used and how patients are using each feature for self-management of their health conditions.

One mechanism by which Web-based medication refills may affect health outcomes may be through improved adherence to prescribed medications. In prior work, MHV use has been associated with improvements in antiretroviral adherence [28]. To the extent that Web-based refills increase the likelihood of refilling prescriptions, they may improve availability of medications, which may lead to improvements in adherence. If the Web-based refill feature improves adherence to antihypertensives and statins, they are likely to improve hypertension and lipid control over time. However, we did not see an association between sustained use of Web-based refills and improvements in HbA1c levels. Because HbA1c is a measure of blood sugar levels over several months, it may take a longer time for improved adherence to diabetes medications to result in measurable improvements, unlike blood pressure and LDL cholesterol, which can result in more rapid improvements even with improved short-term adherence to antihypertensives and statins. A patient's blood sugar levels are also more sensitive to patient diet and self-management, as well as adequate medication titration, both of which may require more patient–provider communication and clinician support to achieve.

SM has been shown to improve patient ratings of patient–provider communication [29]. Thus, SM may also affect adherence by facilitating patient–provider communication about medication or behavioral concerns, which are barriers to adherence. It may also facilitate coordination of care and make it easier for primary care clinicians to refer patients to other related services such as nutrition consults, diabetes counseling, or weight loss programs [30]. This may explain why glycemic control, which requires significant and often complex patient self-management in addition to medication management, was found to be significantly associated with sustained use of SM.

This work also expands on previous research that has often focused generally on the patient portal or PHR use [12] or the use of a particular feature such as SM [9,14,17], without



XSL•FO
RenderX

accounting for their relative effects when used in combination with other features. As features continue to be added to portals, further research should continue to examine the effects of different portal features both separately and in combination, to determine which features are most effective at improving the specific patient outcomes of interest.

Patients who used one or both features during the study period were more likely to be younger, female, white, and were less likely to be socioeconomically disadvantaged than other patients with diabetes who met our inclusion criteria. Numerous studies have documented sociodemographic differences in patient portal access and adoption [26,31-33]. Although we attempted to minimize differences in access by limiting our analyses to patients who had registered for the portal, we still observed differences across groups. It is important to ensure that any improvements in health status achieved through the patient portal do not further widen existing disparities in health because of disparities in portal access or adoption. Lyles et al found that racial or ethnic difference in diabetic patients' shared medical record use was not fully explained by differences in patient sociodemographics, patient health status, or provider encouragement of SM [26]. We will have to be mindful of these potential disparities and specifically target vulnerable patients with support interventions for use of portal features found to significantly affect health outcomes.

### Limitations

There are a number of limitations to this study. The VA patient portal has been deployed nationwide. As all patients are free to choose whether to use the patient portal, it is difficult to limit access or to randomize access to various features to conduct a randomized controlled trial. Because this is an observational study, it is impossible to ensure that the comparison group (ie, the nonusers) is similar in all ways to the portal users. As discussed, we limited the sample to those who had registered to use the portal to reduce heterogeneity in measured and unmeasured confounders. In our prior research [7], we have demonstrated that demographic characteristics were more similar when comparing registered users and nonusers, versus comparing those registered and those not registered. By using patients with diabetes who had registered for MHV (but not used the prescription refill or SM features more than once, if ever,) as a comparison group, we minimized some of this bias by limiting our analyses to patients who had access to the portal and who had attended a training or otherwise shown an interest in using it at some point. We saw that the patients in the comparison group for each logistic regression model were very similar in terms of their baseline health care utilization (number of primary care visits) and number of Elixhauser comorbidities (see Table 2). However, without a measure of patient engagement, there is still the possibility that patients may self-select to use these features precisely because they are already more engaged in their care; the lack of a measure of patient engagement is another limitation of this study. Randomized encouragement trials [34] may be one method to strengthen the rigor of future work.

### Conclusions

Recognizing that our study is an observational study and that the associations cannot be considered causal, the availability of multiple years of observational data, detection of a dose response, and adjustment for patient characteristics known to influence technology use and diabetes outcomes strengthen the potential conclusions we can draw from this analysis about the differential effects use of patient portal features may have on physiological outcomes. The results in this study suggest that measuring the relative use and relative association of each feature of a patient portal is critical because each can have a different effect on changes in health care and health outcomes.

Future research should also focus on uncovering the mechanisms (causal pathways) through which portal use leads to physiological improvements. Does improved communication with providers via SM lead to greater patient engagement between visits, sustained behavior changes, better continuity of care, improved medication titration by the clinical team, or improved adherence to medications by the patients? What portion of the engagement might be explained by other portal features such as the ability to track and chart their blood glucose or blood pressure measurements? A study of adult diabetes patients at Kaiser Permanente found that both patient nonadherence to medications for glycemic, lipid, or blood pressure control and lack of provider treatment intensification occurred frequently among patients whose outcomes are above desired target levels [35]. It may be that portal use assists with patient adherence to medications by facilitating prescription refills, and that patient–provider communication between face-to-face visits can lead to improvements in levels of appropriate treatment intensification by providers. These pathways must be better understood to leverage portal features for interventions.

### Acknowledgments

Dr. Shimada was supported by a Career Development Award (CDA 10-210) from the Department of Veterans Affairs Health Services Research and Development Service and the VA eHealth QUERI (EHQ 10-190). The contents do not represent the views of the US Department of Veterans Affairs or the US Government.

### Conflicts of Interest

None declared.

### References



XSL•FO
RenderX

1. Centers for Disease Control Prevention. Estimates of Diabetes and Its Burden in the United States, 2014. In: National Diabetes Statistics Report. Atlanta, GA: U.S. Department of Health and Human Services; 2014.
2. American DA. Economic costs of diabetes in the U.S. in 2012. Diabetes Care 2013 Apr;36(4):1033-1046. [doi: 10.2337/dc12-2625] [Medline: 23468086]
3. Dall TM, Yang W, Halder P, Pang B, Massoudi M, Wintfeld N, et al. The economic burden of elevated blood glucose levels in 2012: diagnosed and undiagnosed diabetes, gestational diabetes mellitus, and prediabetes. Diabetes Care 2014 Dec;37(12):3172-3179. [doi: 10.2337/dc14-1036] [Medline: 25414388]
4. Nathan DM. Diabetes: Advances in diagnosis and treatment. JAMA 2015 Sep 8;314(10):1052-1062. [doi: 10.1001/jama.2015.9536] [Medline: 26348754]
5. Ali MK, Bullard KM, Saaddine JB, Cowie CC, Imperatore G, Gregg EW. Achievement of goals in U.S. diabetes care, 1999-2010. N Engl J Med 2013 Apr 25;368(17):1613-1624. [doi: 10.1056/NEJMsa1213829] [Medline: 23614587]
6. Gervera K, Graves BA. Integrating Diabetes Guidelines into a Telehealth Screening Tool. Perspect Health Inf Manag 2015;12:1f [FREE Full text] [Medline: 26396557]
7. Shimada SL, Brandt CA, Feng H, McInnes DK, Rao S, Rothendler JA, et al. Personal health record reach in the Veterans Health Administration: a cross-sectional analysis. J Med Internet Res 2014;16(12):e272 [FREE Full text] [doi: 10.2196/jmir.3751] [Medline: 25498515]
8. Institute of Medicine Committee on Quality of Health Care in America. Crossing the quality chasm: a new health system for the 21st century. Washington, DC: National Academy Press; 2001.
9. Weppner WG, Ralston JD, Koepsell TD, Grothaus LC, Reid RJ, Jordan L, et al. Use of a shared medical record with secure messaging by older patients with diabetes. Diabetes Care 2010 Nov;33(11):2314-2319 [FREE Full text] [doi: 10.2337/dc10-1124] [Medline: 20739686]
10. Ricciardi L, Mostashari F, Murphy J, Daniel JG, Siminerio EP. A national action plan to support consumer engagement via e-health. Health Aff (Millwood) 2013 Feb;32(2):376-384 [FREE Full text] [doi: 10.1377/hlthaff.2012.1216] [Medline: 23381531]
11. Tenforde M, Jain A, Hickner J. The value of personal health records for chronic disease management: what do we know? Fam Med 2011 May;43(5):351-354 [FREE Full text] [Medline: 21557106]
12. Tenforde M, Nowacki A, Jain A, Hickner J. The association between personal health record use and diabetes quality measures. J Gen Intern Med 2012 Apr;27(4):420-424 [FREE Full text] [doi: 10.1007/s11606-011-1889-0] [Medline: 22005937]
13. Wade-Vuturo A, Mayberry LS, Osborn CY. Secure messaging and diabetes management: experiences and perspectives of patient portal users. J Am Med Inform Assoc 2013 May 1;20(3):519-525 [FREE Full text] [doi: 10.1136/amiajnl-2012-001253] [Medline: 23242764]
14. 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 [FREE Full text] [doi: 10.1377/hlthaff.2010.0048] [Medline: 20606190]
15. Harris L, Haneuse SJ, Martin DP, Ralston JD. Diabetes quality of care and outpatient utilization associated with electronic patient-provider messaging: a cross-sectional analysis. Diabetes Care 2009 Jul;32(7):1182-1187 [FREE Full text] [doi: 10.2337/dc08-1771] [Medline: 19366959]
16. Harris L, Koepsell TD, Haneuse SJ, Martin DP, Ralston JD. Glycemic control associated with secure patient-provider messaging within a shared electronic medical record: a longitudinal analysis. Diabetes Care 2013 Sep;36(9):2726-2733 [FREE Full text] [doi: 10.2337/dc12-2003] [Medline: 23628618]
17. Sarkar U, Lyles CR, Parker MM, Allen J, Nguyen R, Moffet HH, et al. Use of the refill function through an online patient portal is associated with improved adherence to statins in an integrated health system. Med Care 2014 Mar;52(3):194-201 [FREE Full text] [doi: 10.1097/MLR.0000000000000069] [Medline: 24374412]
18. Kaelber D, Pan EC. The value of personal health record (PHR) systems. AMIA Annual Symposium Proceedings 2008;Nov 6:343-347. [Medline: 18999276]
19. 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]
20. Nazi K, Woods SS. MyHealtheVet PHR: a description of users and patient portal use. AMIA Annu Symp Proc 2008;Nov 6:1182. [Medline: 18999142]
21. Chumbler NR, Haggstrom D, Saleem JJ. Implementation of health information technology in Veterans Health Administration to support transformational change: telehealth and personal health records. Med Care 2011 Dec;49 Suppl:S36-S42. [doi: 10.1097/MLR.0b013e3181d558f9] [Medline: 20421829]
22. 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]
23. 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;17(2):203-211 [FREE Full text] [doi: 10.1136/jamia.2009.000240] [Medline: 20190065]




XSL•FO
RenderX

24. Miller DR, Safford MM, Pogach LM. Who has diabetes? Best estimates of diabetes prevalence in the Department of Veterans Affairs based on computerized patient data. Diabetes Care 2004 May;27 Suppl 2:B10-B21. [Medline: 15113777]
25. American Diabetes Association. Standards of Medical Care in Diabetes—2013. Diabetes Care 2013;36(Suppl 1):S11-S66 [FREE Full text] [doi: 10.2337/dc13-S011]
26. Lyles CR, Harris LT, Jordan L, Grothaus L, Wehnes L, Reid RJ, et al. Patient race/ethnicity and shared medical record use among diabetes patients. Med Care 2012 May;50(5):434-440. [doi: 10.1097/MLR.0b013e318249d81b] [Medline: 22354209]
27. Quan H, Sundararajan V, Halfon P, Fong A, Burnand B, Luthi J, et al. Coding algorithms for defining comorbidities in ICD-9-CM and ICD-10 administrative data. Med Care 2005 Nov;43(11):1130-1139. [Medline: 16224307]
28. McInnes DK, Shimada SL, Rao SR, Quill A, Duggal M, Gifford AL, et al. Personal health record use and its association with antiretroviral adherence: survey and medical record data from 1871 US veterans infected with HIV. AIDS Behav 2013 Nov;17(9):3091-3100. [doi: 10.1007/s10461-012-0399-3] [Medline: 23334359]
29. Lin C, Wittevrongel L, Moore L, Beaty BL, Ross SE. An Internet-based patient-provider communication system: randomized controlled trial. J Med Internet Res 2005;7(4):e47 [FREE Full text] [doi: 10.2196/jmir.7.4.e47] [Medline: 16236699]
30. 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;11(4):260-267 [FREE Full text] [doi: 10.1197/jamia.M1445] [Medline: 15064295]
31. Smith SG, O'Conor R, Aitken W, Curtis LM, Wolf MS, Goel MS. Disparities in registration and use of an online patient portal among older adults: findings from the LitCog cohort. J Am Med Inform Assoc 2015 Jul;22(4):888-895 [FREE Full text] [doi: 10.1093/jamia/ocv025] [Medline: 25914099]
32. Roblin DW, Houston TK, Allison JJ, Joski PJ, Becker ER. Disparities in use of a personal health record in a managed care organization. J Am Med Inform Assoc 2009;16(5):683-689 [FREE Full text] [doi: 10.1197/jamia.M3169] [Medline: 19567790]
33. Ralston J, Rutter CM, Carrell D, Hecht J, Rubanowice D, Simon GE. Patient use of secure electronic messaging within a shared medical record: a cross-sectional study. J Gen Intern Med 2009 Mar;24(3):349-355 [FREE Full text] [doi: 10.1007/s11606-008-0899-z] [Medline: 19137379]
34. Connell AM. Employing complier average causal effect analytic methods to examine effects of randomized encouragement trials. Am J Drug Alcohol Abuse 2009;35(4):253-259 [FREE Full text] [doi: 10.1080/00952990903005882] [Medline: 20180678]
35. Schmittdiel J, Uratsu CS, Karter AJ, Heisler M, Subramanian U, Mangione CM, et al. Why don't diabetes patients achieve recommended risk factor targets? Poor adherence versus lack of treatment intensification. J Gen Intern Med 2008 May;23(5):588-594 [FREE Full text] [doi: 10.1007/s11606-008-0554-8] [Medline: 18317847]


## Abbreviations

**BP:** blood pressure
**HbA1c:** hemoglobin A1c
**ICD-9-CM:** International Classification of Diseases, Ninth Revision, Clinical Modification
**LDL:** low density lipoprotein
**MHV:** My HealtheVet
**PHR:** personal health record
**SM:** secure messaging
**VA:** Department of Veterans Affairs


*Edited by G Eysenbach; submitted 22.02.16; peer-reviewed by K Blondon, J Whealin; comments to author 17.03.16; revised version received 03.05.16; accepted 19.05.16; published 01.07.16.*

*Please cite as:*
*Shimada SL, Allison JJ, Rosen AK, Feng H, Houston TK*
*Sustained Use of Patient Portal Features and Improvements in Diabetes Physiological Measures*
*J Med Internet Res 2016;18(7):e179*
*URL: http://www.jmir.org/2016/7/e179/*
*doi:10.2196/jmir.5663*
*PMID:27369696*


©Stephanie L Shimada, Jeroan J Allison, Amy K Rosen, Hua Feng, Thomas K Houston. Originally published in the Journal of Medical Internet Research (http://www.jmir.org), 01.07.2016. This is an open-access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/2.0/), which permits unrestricted use, distribution,



XSL•FO
RenderX

and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on http://www.jmir.org/, as well as this copyright and license information must be included.



XSL•FO
RenderX

Original Paper

# The Effect of a Freely Available Flipped Classroom Course on Health Care Worker Patient Safety Culture: A Prospective Controlled Study

Lowell Ling[1], MBBS MPhil; Charles David Gomersall[2], MBBS, FCICM; Winnie Samy[2], BN, MSc; Gavin Matthew Joynt[2], MB BCh, FCICM; Czarina CH Leung[2], MB ChB, FHKAM; Wai-Tat Wong[2], MBBS, FHKAM; Anna Lee[2], MPH, PhD

[1]Prince of Wales Hospital, Department of Anaesthesia and Intensive Care, Shatin, China (Hong Kong)

[2]The Chinese University of Hong Kong, Department of Anaesthesia and Intensive Care, Shatin, China (Hong Kong)

**Corresponding Author:**
Anna Lee, MPH, PhD
The Chinese University of Hong Kong
Department of Anaesthesia and Intensive Care
4th Floor, Main Clinical Block and Trauma Centre
Prince of Wales Hospital
Shatin,
China (Hong Kong)
Phone: 852 2632 2735
Fax: 852 2637 2422
Email: annalee@cuhk.edu.hk

## Abstract

**Background:** Patient safety culture is an integral aspect of good standard of care. A good patient safety culture is believed to be a prerequisite for safe medical care. However, there is little evidence on whether general education can enhance patient safety culture.

**Objective:** Our aim was to assess the impact of a standardized patient safety course on health care worker patient safety culture.

**Methods:** Health care workers from Intensive Care Units (ICU) at two hospitals (A and B) in Hong Kong were recruited to compare the changes in safety culture before and after a patient safety course. The BASIC Patient Safety course was administered only to staff from Hospital A ICU. Safety culture was assessed in both units at two time points, one before and one after the course, by using the Hospital Survey on Patient Safety Culture questionnaire. Responses were coded according to the Survey User's Guide, and positive response percentages for each patient safety domain were compared to the 2012 Agency for Healthcare Research and Quality ICU sample of 36,120 respondents.

**Results:** We distributed 127 questionnaires across the two hospitals with an overall response rate of 74.8% (95 respondents). After the safety course, ICU A significantly improved on teamwork within hospital units ($P$=.008) and hospital management support for patient safety ($P$<.001), but decreased in the frequency of reporting mistakes compared to the initial survey ($P$=.006). Overall, ICU A staff showed significantly greater enhancement in positive responses in five domains than staff from ICU B. Pooled data indicated that patient safety culture was poorer in the two ICUs than the average ICU in the Agency for Healthcare Research and Quality database, both overall and in every individual domain except hospital management support for patient safety and hospital handoffs and transitions.

**Conclusions:** Our study demonstrates that a structured, reproducible short course on patient safety may be associated with an enhancement in several domains in ICU patient safety culture.

***(J Med Internet Res 2016;18(7):e180)*** doi:10.2196/jmir.5378

**KEYWORDS**

patient safety; critical care; education, professional; education, distance; safety culture



XSL•FO
RenderX

## Introduction

Good medical practice is based on the classic maxim of "primum non nocere," and yet it is estimated that at least 1 in 10 patients may be harmed by adverse events during their hospital stay [1,2]. The landmark US Institute of Medicine's report in 1999, *To Err is Human: Building a Safer Health System*, ignited the interest in improving patient safety. Studies into harm suggest that a significant proportion of adverse events is preventable [3]. These include prescription mistakes, handover lapses, surgical errors, diagnostic mishaps, and other errors attributable to the human factor [4-7].

Changing and adopting health care technology has been shown to reduce medical errors and improve patient safety [8]. However, technology itself improves patient safety only to a limited extent, and further error reduction requires human factors and organizational change [9]. Clinical human interventions such as additional pharmacist inspection of electronic prescriptions can further reduce medication errors [6]. Using targeted education to change clinical practice seems to be effective as well. Specific interventions adopted for central venous catheter insertions have been shown to reduce central venous access-related infection and improve patient outcome [10]. However, it has also been shown that although targeted educational interventions could improve clinical staff knowledge, this did not translate to improved outcomes [11]. This highlights multiple challenges in studying the effect of education. First, knowledge itself may be a prerequisite for safety culture. However, attitudes and perception are equally important but more difficult to measure and define. Second, education is often not standardized, as such the findings may not be generalizable. Third, it is difficult to conduct blinded randomized trials with appropriate controls. Studies on patient safety education programs have generally not assessed the effectiveness of these interventions with adequate controls and rigor [12].

Recent reports on the failings of a hospital in the United Kingdom have highlighted issues with a lack of an appropriate patient safety culture [13]. The likely causal relationship between poor culture and poor patient care stresses the importance of improving culture to improve standards of care. Therefore, we decided to conduct a prospective controlled study to assess the impact of a standardized, free license, patient safety course on patient safety culture.

## Methods

### Study Design and Hospitals

The study protocol was approved by the Chinese University of Hong Kong Survey and Behavioural Research Ethics Committee. This was a prospective controlled, before and after, study design that used the Hospital Survey on Patient Safety Culture (HSOPSC) questionnaire instrument to evaluate the impact of the BASIC (Basic Assessment and Support in Intensive Care) Patient Safety Course on safety culture. The course was delivered to doctors, nurses, and health care assistants in the Intensive Care Unit (ICU) of hospital A (ICU A) only. In order to control for temporal changes or changes resulting from policies implemented across the entire public hospital system, the questionnaire was also administered to equivalent staff in the ICU of hospital B, a neighboring hospital (ICU B). The two hospitals are publicly funded, located in the same hospital cluster, and share a common cluster chief executive. Hospital A is a tertiary teaching hospital with 1400 beds and 22 ICU beds. Hospital B is an acute general hospital with 600 inpatient beds and 14 ICU beds. A comparison of clinical data between ICU A and ICU B is shown in Table 1.

**Table 1.** Comparison Statistics of ICU A and B.

| | Admissions per year | Severity of illness (APACHE III Acute Physiology Score) | Average ICU length of stay | Risk-adjusted hospital mortality ratio |
|---|---|---|---|---|
| ICU A | 1500 | 50 | 4 days | 0.80 |
| ICU B | 600 | 55 | 5 days | 0.75 |

### BASIC Patient Safety Course

The BASIC Patient Safety course is a blended learning course that uses a flipped classroom approach in which didactic teaching is carried out prior to participants attending face-to-face teaching. This allows face-to-face time to be dedicated to interactive sessions involving application of the knowledge already acquired. In our course, preparatory material consists of a short printed course manual and e-learning.

The e-learning material comprises short narrated lectures, typically based around a modified clinical case, formative assessment, an interactive electronic lesson, and a video of an incident involving a serious medication error. The e-lectures were created in PowerPoint and Camtasia and were produced as MP4 files so that they could be played on different platforms including Windows, iOS, and Android. The files were uploaded to a Moodle 2.0 platform that was configured to be accessible both on personal computers and mobile devices. Each lecture is supplemented by a formative assessment, in the form of a multiple choice test, which emphasizes the key points covered in the lecture. The system is configured to allow participants to access the assessment only after watching the corresponding lecture for its entire duration.

The interactive lesson is a more complex form of formative assessment. Each participant's individual pathway through the lesson is dependent on their answers to questions posed during the lesson. Candidates with a poorer understanding of the material take a longer pathway, receiving greater explanation of basic aspects before moving on to more complex issues, while those with a greater understanding rapidly progress to the more complex material. Teaching is thus adjusted to the participant's needs. All activity on the e-learning site is



XSL•FO
RenderX

automatically logged (with the knowledge of participants). Completion of all e-learning is required before attending the face-to-face teaching. Computers are made available to participants who are unable to access the material on their own devices, but no participants required this facility during the study.

Face-to-face teaching consists of small group teaching involving a simulated emergency (to practice communication and team and leadership skills), practicing Situation Background Assessment Recommendation (SBAR) communication, practicing breaking news of an error, discussion of the video shown on the e-learning site, and reflection on patient safety in one's own unit. This was followed by a group debriefing session. The face-to-face component of the course lasts 3 hours.

The course was written specifically for health care professionals whose primary function is to provide clinical care. It is not aimed at those with a predominantly managerial role. Participants are expected to gain detailed knowledge of the definitions and scope of patient safety, human factors engineering and why it is important to patient safety; cause and reduction of errors, preventing errors leading to harm, cognition, communication with colleagues and patients and the importance of full disclosure, root cause analysis, quality improvement, teamwork, medication safety, and coping with errors. They are expected to enhance the skills required to be an effective team player, understand and learn from errors, understand and manage clinical risk, engage with patients and caregivers, and communicate with full disclosure after adverse events. The course material is available free of charge, and the course is disseminated on a train-the-trainers basis. Facilitators/instructors for the small group teaching are given detailed written guidance on the content for each discussion/tutorial. One instructor is required for every 6 participants.

Between April to December 2011, 117 participants attended the course, of whom 91 nurses and 8 doctors worked in ICU A. No staff from ICU B attended the course. The course was taught predominantly by senior nursing and medical staff from ICU A.

### Measurements

A convenience sample of doctors, nurses, and health care assistants in the ICU from the two hospitals were asked to complete a Hong Kong Chinese version of the HSOPSC before and after the course was implemented to measure their attitudes towards patient safety. The survey consisted of 44 questions with 5-point Likert response scale of agreement of either strongly disagree to strongly agree, never to always, or failing to excellent. Our version contained an extra question about whether the survey participant would feel safe being treated as a patient in the respective hospital. These questions were grouped into 13 dimensions and the overall positive scores were calculated by categorizing strongly agree/agree, excellent/very good, or always/most of time as positive responses. For questions that were negatively worded, positive responses equated to strongly disagree/disagree or never/rarely. The survey also asked the frequency of event reporting by the participant over the past year.

The pre-course (baseline) survey was carried out immediately before the first time the course was run, and the post course survey within 3 months of completion of the series of courses. We wanted to assess the effects of the course on general staff patient safety culture rather than specifically on the attitudes of course attendees. Therefore, the survey respondents were selected randomly from the staff of ICU A and ICU B. Course feedback was collected from participants using anonymized electronic feedback forms.

### Statistical Analyses

We entered the responses into the AHRQ Hospital Survey on Patient Safety Culture Excel tool (version 1.5). Positive responses were coded according to the Survey User's Guide. For each hospital, the percentage change was estimated as the follow-up percentage minus the baseline percentage. The 95% confidence interval (95% CI) around the positive response percentage for each AHQR patient safety domain was estimated, and baseline results were compared to the 2012 AHRQ ICU sample of 36,120 respondents to provide a contextual reference for interpretation of the applicability of our findings.

Separate generalized estimating equations (GEE) were used to account for the correlation in participants responding to both pre- and post-workshop questionnaires. Separate difference-in-differences models were constructed for each of the 12 AHRQ patient safety domains using participant-level data [14]. The outcome "positive response" was modelled as a function of ICU (A or B), period (baseline or post-intervention), and an interaction term between ICU and period, adjusted for duration of work in Intensive Care (≤10 years vs >10 years). The coefficient for the interaction term in the GEE model indicated whether ICU A improved more or less than ICU B from baseline to follow up. All statistical analysis was carried out using SPSS version 22.0.

## Results

### Response Rates and Participant Characteristics

The total number of questionnaires distributed across the two hospitals was 127 with an overall response rate of 74.8% (95 respondents). Three respondents answered both pre- and post-workshop questionnaires. The pre- and post-intervention response rates from ICU A were 88% (37/42) and 79% (23/29), respectively. The response rates from ICU B were 63% for both pre- (20/32) and post-intervention (15/24) survey.

Of the 95 participants, 78 were registered ICU nurses, 11 patient care assistants, and 6 physicians. Most respondents (90/95, 95%) had direct contact or interactions with patients. Over half (53/95, 56%) had worked in the hospital system for less than 10 years. There was no difference in the proportion of participants working less than or equal to 10 years in the current work area/unit between hospitals: ICU A 58% (33/57) versus ICU B 74% (28/38), $P$=.12.

### Survey Responses

ICU A had lower positive responses at baseline on 7 of 12 domains when compared to baseline responses of ICU B (see Table 2). These included overall perception of safety ($P$=.007),



XSL•FO
RenderX

organization learning/continuous improvement ($P$=.03), teamwork within hospital units ($P$=.003), communication openness ($P$=.03), feedback and communication about error ($P$<.001), staffing ($P$<.001), and hospital management support for patient safety ($P$=.02).

After the safety course, ICU A had significantly improved responses in teamwork within hospital units ($P$=.008) and hospital management support for patient safety ($P$<.001), but decreased in the frequency of reporting mistakes ($P$=.006) compared to baseline. For ICU B, there was a decrease in the proportion of positive responses in 6 measured domains during the same period (see Table 2).

There was a significant interaction between ICUs and period, after adjusting for the duration of work in current area/unit, indicating that ICU A showed greater improvement in positive responses than ICU B in 5 domains (see Table 3).

**Table 2.** Unadjusted difference in positive responses at baseline and follow-up between hospitals.

| Domain | Baseline responses (%) | | Follow-up responses (%) | | Changes from baseline (%)[a] | | | |
|---|---|---|---|---|---|---|---|---|
| | ICU B | ICU A | ICU B | ICU A | ICU B | *P* | ICU A | *P* |
| Frequency of reporting | 32/60 (53.3) | 43/108 (39.8) | 14/43 (32.6) | 13/66 (19.7) | -20.8 | .04 | -20.1 | .006 |
| Overall perception of safety | 39/80 (48.8) | 42/148 (28.3) | 21/60 (35.0) | 37/92 (40.2) | -13.8 | .10 | 11.9 | .06 |
| Supervisor/manager expectations and actions promoting safety | 55/80 (68.8) | 82/148 (55.4) | 33/60 (55.0) | 60/92 (65.2) | -13.8 | .10 | 9.8 | .13 |
| Organization learning/continuous improvement | 45/60 (75.0) | 61/111 (55.0) | 27/45 (60.0) | 44/69 (63.8) | -15.0 | .10 | 8.8 | .24 |
| Teamwork within hospital units | 71/80 (88.8) | 91/147 (61.9) | 40/60 (66.7) | 72/92 (78.3) | -22.1 | .001 | 16.4 | .008 |
| Communication openness | 31/60 (51.7) | 34/111 (30.6) | 12/45 (26.7) | 21/69 (30.4) | -25.0 | .01 | -0.2 | .98 |
| Feedback and communication about error | 47/60 (78.3) | 40/111 (36.0) | 18/45 (40.0) | 34/69 (49.3) | -38.3 | <.001 | 13.3 | .08 |
| Nonpunitive response | 19/60 (31.7) | 19/111 (17.1) | 6/45 (13.3) | 10/69 (14.5) | -18.4 | .03 | -2.6 | .64 |
| Staffing | 48/80 (60.0) | 42/147 (28.6) | 24/60 (40.0) | 32/90 (35.6) | -20.0 | .02 | 7.0 | .26 |
| Hospital management support for patient safety | 39/60 (65.0) | 44/108 (40.7) | 27/45 (60.0) | 48/69 (69.6) | -5.0 | .60 | 28.9 | <.001 |
| Teamwork across hospital units | 41/80 (51.3) | 68/144 (47.2) | 23/60 (38.3) | 39/92 (42.4) | -13.0 | .13 | -4.8 | .47 |
| Hospital handoffs and transitions | 40/79 (50.6) | 70/143 (49.0) | 34/59 (57.6) | 43/92 (46.7) | -7.0 | .42 | -2.3 | .74 |

[a]Follow-up percentage minus the baseline percentage. Denominators for each item are the product of the number of questions in that domain and the number of respondents. Numerators are the total number of positive responses to all questions in that domain.



XSL•FO
RenderX

**Table 3.** Relative risk (95% CI) of improvement in patient safety domains: Baseline to follow-up in hospitals with and without educational intervention.

| Domain | Relative risk (95% CI for difference between groups)[a] | *P* value |
|---|---|---|
| Frequency of reporting | 0.90 (0.33-2.49) | .84 |
| Overall perception of safety | 1.94 (1.11-3.37) | .02 |
| Supervisor/manager expectations and actions promoting safety | 1.48 (0.99-2.20) | .06 |
| Organization learning/continuous improvement | 1.45 (0.96-2.20) | .08 |
| Teamwork within hospital units | 1.55 (1.10-2.19) | .01 |
| Communication openness | 1.66 (0.73-3.76) | .23 |
| Feedback and communication about error | 2.47 (1.28-4.80) | .007 |
| Nonpunitive response | 1.68 (0.54-5.18) | .37 |
| Staffing | 1.92 (1.15-3.19) | .01 |
| Hospital management support for patient safety | 1.88 (1.16-3.04) | .01 |
| Teamwork across hospital units | 1.23 (0.75-2.00) | .41 |
| Hospital handoffs and transitions | 0.86 (0.44-1.70) | .67 |

[a]Adjusted for duration of work in current area/unit (≤10 years vs >10 years)

Participants in ICU A were seven times more likely to report "feeling safe being treated in this hospital as a patient" than those in ICU B after adjusting for duration of work in the current area/unit (*P*=.01; see Table 4).

**Table 4.** Response to statement "I would feel safe being treated in this hospital as a patient."

| | Baseline responses (%) | | Follow-up responses (%) | | Changes from baseline responses (%) | | |
|---|---|---|---|---|---|---|---|
| | ICU B | ICU A | ICU B | ICU A | ICU B | ICU A | Relative risk[a] (95% CI) |
| Feel safe | 11/19 (57.9) | 14/35 (40.0) | 3/15 (20.0) | 12/22 (54.5) | -37.9 | 14.5 | 7.29 (1.52-34.94) |

[a]Interaction effect (risk ratio of improvement from base to follow-up between ICUs, adjusted for duration of work in current area/unit (≤10 years vs >10 years).

Pooled data of all participants from both ICUs indicate that patient safety culture was poorer than the average ICUs in the 2012 AHRQ database (see Table 5). Reponses from both ICUs were lower (at least 5% point difference) for every individual domain except hospital management support for patient safety and hospital handoffs and transitions.

**Table 5.** Domain-level comparative average percentage (95% CI) positive responses of Hong Kong ICUs (N=95) to 2012 AHRQ database (N=36,120).

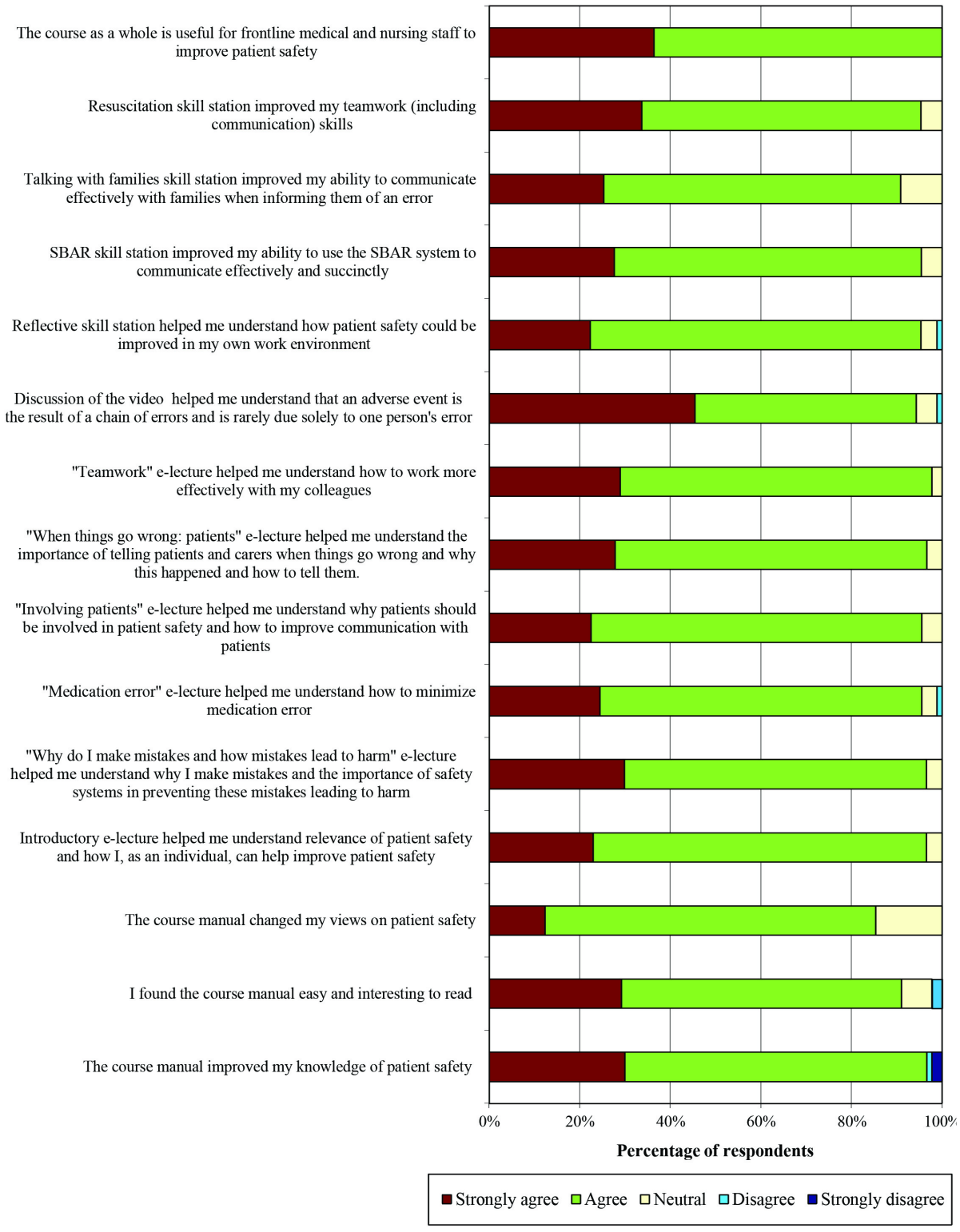
| Domain | Hong Kong ICUs, % (95% CI) | 2012 AHRQ ICUs, % |
|---|---|---|
| Frequency of reporting | 37 (28-47) | 59 |
| Overall perception of safety | 37 (28-47) | 60 |
| Supervisor/manager expectations and actions promoting safety | 61 (51-70) | 73 |
| Organization learning/continuous improvement | 62 (52-71) | 72 |
| Teamwork within hospital units | 72 (62-80) | 84 |
| Communication openness | 34 (25-44) | 61 |
| Feedback and communication about error | 49 (40-59) | 60 |
| Nonpunitive response | 19 (12-28) | 40 |
| Staffing | 39 (30-49) | 58 |
| Hospital management support for patient safety | 56 (46-65) | 64 |
| Teamwork across hospital units | 45 (36-55) | 57 |
| Hospital handoffs and transitions | 50 (39-59) | 51 |
| Average across domains | 47 (38-57) | 62 |



XSL•FO
RenderX

Of the 117 safety course participants, 90 (77%) answered the course feedback questionnaire. The vast majority of these 90 participants agreed or strongly agreed with positive statements about the course (see Figure 1), with 32% strongly agreeing and another 56% agreeing that the course was useful to improve patient safety. Notably, participants were as likely to agree or strongly agree with positive statements about the electronic lectures as they were to agree or strongly agree with positive statements about other aspects of the course.

**Figure 1.** Participants responses to the feedback questionnaire.






XSL•FO
RenderX

## Discussion

### Principal Findings

Our data suggest that a structured, reproducible, short blended learning course on patient safety may improve perceived ICU patient safety culture. After controlling for duration of working in the respective ICUs, there was a significant improvement in 5 of 12 domains and a trend towards improvement in 2 others in the ICU where the course was given. Furthermore there was a substantial difference in the change of response to the additional statement "I would feel safe being treated in this hospital as a patient" in favor of ICU A.

However, it is notable that there was a deterioration in patient safety culture in ICU B during the study period. This could be due to factors unique to that ICU or hospital or to systemic changes that would also have affected ICU A. If the decline in ICU B was due to systemic factors, then our results suggest that our course not only arrested but largely reversed the deterioration. On the other hand, if the deterioration was due to unique factors affecting only ICU B, then the positive effect of our course could be less than our results would suggest. Nevertheless, if one compares only pre- and post-course results in ICU A in isolation (see Table 2), it can be seen that there were still significant improvements in the domains of teamwork within hospital units and hospital management support for patient safety, and weak evidence supporting an improvement in the feedback and communication about error and overall perception of safety domains.

There was a significant deterioration in the frequency of reporting domain in both ICUs during the study period. In ICU B, this may simply reflect the general deterioration seen across multiple domains. In ICU A, the change is difficult to explain as it is the only domain in which there was a significant reduction. One possibility is that the changes in ICU B did reflect a systemic deterioration in patient safety culture across the two hospitals that was reversed in most, but not all, domains in ICU A by the course. However, we cannot exclude the possibility that this was an inadvertent adverse effect of the course.

### Limitations

Our study has a number of other weaknesses. Similar to all before and after studies, we cannot exclude the possibility of confounding factors that affected only ICU A. Rather than a direct comparison between two ICUs, this study evaluated the temporal change in each unit. Therefore the relevant confounding factors are ones that affect temporal changes rather than the baseline differences between the two ICUs. What we have shown is a temporal relationship between our course and changes in patient safety attitudes, not a causal relationship. The feedback data suggest that the course may have changed attitudes. However, it is possible that the responses of some of the participants from ICU A may have been influenced by the fact that the course was taught by senior staff from the same ICU, even though the feedback was anonymous. Furthermore, any change in attitude may have been due to the involvement of senior staff signaling to other staff that patient safety is an important issue rather than the educational content of the course itself.

We studied only two ICUs; therefore, our results may not be generalizable to other ICUs or to other hospital departments. Previous studies on patient safety culture in Chinese countries such as China and Taiwan showed some important ethnic and cultural factors that may result in differences to western patient safety culture [15-18]. Furthermore, the baseline data suggest that the patient safety culture was poor in both ICUs, relative to ICUs contributing to the AHRQ database, and it is possible that the course may have little effect in units where patient safety culture is well developed. Finally, we have studied only the short-term effect of the course, and although appropriate patient safety culture is considered a pre-requisite to patient safety behavior in practice, this has not been rigorously tested [19].

### Study Strengths

Our study does have the advantage that we studied the effect of a standardized, freely available, educational intervention with a parallel control group [14,20]. The standardized nature of the course both facilitates further research (ie, the same intervention is tested each time) and the applicability of the results (course material may be obtained directly from the authors). Studies of educational interventions that are not in the public domain cannot be reproduced, and it is unclear whether the results can be applied to other individual educational packages. Previous studies of educational interventions to improve patient safety show variable results suggesting that the exact nature of the intervention may be important [10,11,21-24]. In particular, our course incorporates e-learning, interactive modules, and formative assessments. Active involvement and formative assessment are key elements for effective adult learning. A similar educational approach has been used by the Canadian "Managing Obstetric Risks Efficiently" safety program and proven to be effective in advancing safety knowledge (culture was not examined) [25].

Although the course is labor-intensive with a high ratio of instructors to participants, the face-to-face contact time is short as a result of the pre-course reading and e-learning. This facilitates its use as part of in-service training, by minimizing disruption to clinical staffing. It is notable that the e-learning components of the course were highly rated by participants. Although we did not test the specific effect of the e-learning, a study of one of our other courses suggests it enhances learning [26]. The same study revealed that participants value the flexibility of listening to e-lectures at their own convenience and the ability to re-play lectures.

### Conclusions

Our results suggest that the course may improve patient safety culture. Further research is required to establish whether the temporal association can be reproduced when more units are studied in a variety of different cultures and if so, an attempt should be made to determine whether the relationship is causal. In conclusion, introduction of a standardized patient safety course was temporally associated with an improvement in several domains of patient safety culture in a single ICU.



XSL•FO
RenderX

## Acknowledgments

We thank the Hong Kong Hospital Authority and the Chinese University of Hong Kong Grants Committee Knowledge Transfer Project Fund (KTP091126F06) for funding and support.

## Authors' Contributions

LLWL drafted the manuscript and made substantial revisions. AL performed the statistical analyses and had full access to all the data in the study. CDG, GMJ, and AL were involved in the study concept and design of the study. WS collected the data. All authors interpreted the data, made critical revisions to the manuscript for important intellectual content, and approved the final version of the manuscript.

## Conflicts of Interest

CDG and GMJ were involved in developing and teaching the BASIC course.

## References

1. Davis P, Lay-Yee R, Briant R, Ali W, Scott A, Schug S. Adverse events in New Zealand public hospitals I: occurrence and impact. N Z Med J 2002 Dec 13;115(1167):U271. [Medline: 12552260]
2. de Vries EN, Ramrattan MA, Smorenburg SM, Gouma DJ, Boermeester MA. The incidence and nature of in-hospital adverse events: a systematic review. Qual Saf Health Care 2008 Jun;17(3):216-223 [FREE Full text] [doi: 10.1136/qshc.2007.023622] [Medline: 18519629]
3. Vincent C, Neale G, Woloshynowych M. Adverse events in British hospitals: preliminary retrospective record review. BMJ 2001 Mar 3;322(7285):517-519 [FREE Full text] [Medline: 11230064]
4. Leape LL, Brennan TA, Laird N, Lawthers AG, Localio AR, Barnes BA, et al. The nature of adverse events in hospitalized patients. Results of the Harvard Medical Practice Study II. N Engl J Med 1991 Feb 7;324(6):377-384. [doi: 10.1056/NEJM199102073240605] [Medline: 1824793]
5. Thomas EJ, Studdert DM, Burstin HR, Orav EJ, Zeena T, Williams EJ, et al. Incidence and types of adverse events and negligent care in Utah and Colorado. Med Care 2000 Mar;38(3):261-271. [Medline: 10718351]
6. Abdel-Qader DH, Harper L, Cantrill JA, Tully MP. Pharmacists' interventions in prescribing errors at hospital discharge: an observational study in the context of an electronic prescribing system in a UK teaching hospital. Drug Saf 2010 Nov 1;33(11):1027-1044. [doi: 10.2165/11538310-000000000-00000] [Medline: 20925440]
7. Starmer AJ, Sectish TC, Simon DW, Keohane C, McSweeney ME, Chung EY, et al. Rates of medical errors and preventable adverse events among hospitalized children following implementation of a resident handoff bundle. JAMA 2013 Dec 4;310(21):2262-2270. [doi: 10.1001/jama.2013.281961] [Medline: 24302089]
8. Ball MJ, Garets DE, Handler TJ. Leveraging information technology towards enhancing patient care and a culture of safety in the US. Methods Inf Med 2003;42(5):503-508. [doi: 10.1267/METH03050503] [Medline: 14654884]
9. Anderson JG, Ramanujam R, Hensel D, Anderson MM, Sirio CA. The need for organizational change in patient safety initiatives. Int J Med Inform 2006 Dec;75(12):809-817. [doi: 10.1016/j.ijmedinf.2006.05.043] [Medline: 16870501]
10. Bion J, Richardson A, Hibbert P, Beer J, Abrusci T, McCutcheon M, Matching Michigan Collaboration & Writing Committee. 'Matching Michigan': a 2-year stepped interventional programme to minimise central venous catheter-blood stream infections in intensive care units in England. BMJ Qual Saf 2013 Feb;22(2):110-123 [FREE Full text] [doi: 10.1136/bmjqs-2012-001325] [Medline: 22996571]
11. Santana SL, Furtado GH, Wey SB, Medeiros EA. Impact of an education program on the incidence of central line-associated bloodstream infection in 2 medical-surgical intensive care units in Brazil. Infect Control Hosp Epidemiol 2008 Dec;29(12):1171-1173. [doi: 10.1086/591862] [Medline: 18956978]
12. Halligan M, Zecevic A. Safety culture in healthcare: a review of concepts, dimensions, measures and progress. BMJ Qual Saf 2011 Apr;20(4):338-343. [doi: 10.1136/bmjqs.2010.040964] [Medline: 21303770]
13. Investigation into mid Staffordshire NHS Foundation Trust. United Kingdom: Healthcare Commission; 2009 Mar. URL: http://www.nhshistory.net/midstaffs.pdf [accessed 2015-10-16] [WebCite Cache ID 6iKQaEMpq]
14. Dimick JB, Ryan AM. Methods for evaluating changes in health care policy: the difference-in-differences approach. JAMA 2014 Dec 10;312(22):2401-2402. [doi: 10.1001/jama.2014.16153] [Medline: 25490331]
15. Liu Y, Kalisch BJ, Zhang L, Xu J. Perception of safety culture by nurses in hospitals in china. J Nurs Care Qual 2009;24(1):63-68. [doi: 10.1097/NCQ.0b013e31818f551f] [Medline: 19092481]
16. Nie Y, Mao X, Cui H, He S, Li J, Zhang M. Hospital survey on patient safety culture in China. BMC Health Serv Res 2013;13:228 [FREE Full text] [doi: 10.1186/1472-6963-13-228] [Medline: 23800307]
17. Chen I, Li HH. Measuring patient safety culture in Taiwan using the Hospital Survey on Patient Safety Culture (HSOPSC). BMC Health Serv Res 2010;10:152 [FREE Full text] [doi: 10.1186/1472-6963-10-152] [Medline: 20529246]



XSL•FO
RenderX

18. Zhu J, Li L, Li Y, Shi M, Lu H, Garnick DW, et al. What constitutes patient safety culture in Chinese hospitals? Int J Qual Health Care 2012 Jun;24(3):250-257 [FREE Full text] [doi: 10.1093/intqhc/mzs010] [Medline: 22461205]
19. Wakefield JG, McLaws ML, Whitby M, Patton L. Patient safety culture: factors that influence clinician involvement in patient safety behaviours. Qual Saf Health Care 2010 Dec;19(6):585-591. [doi: 10.1136/qshc.2008.030700] [Medline: 20724390]
20. Villa JM. diff: Simplifying the estimation of difference-in-differences treatment effects. The Stata Journal 2016 Mar;16(1):52-71.
21. Warren DK, Zack JE, Cox MJ, Cohen MM, Fraser VJ. An educational intervention to prevent catheter-associated bloodstream infections in a nonteaching, community medical center. Crit Care Med 2003 Jul;31(7):1959-1963. [doi: 10.1097/01.CCM.0000069513.15417.1C] [Medline: 12847389]
22. Warren DK, Zack JE, Mayfield JL, Chen A, Prentice D, Fraser VJ, et al. The effect of an education program on the incidence of central venous catheter-associated bloodstream infection in a medical ICU. Chest 2004 Nov;126(5):1612-1618. [doi: 10.1378/chest.126.5.1612] [Medline: 15539735]
23. Sherertz RJ, Ely EW, Westbrook DM, Gledhill KS, Streed SA, Kiger B, et al. Education of physicians-in-training can decrease the risk for vascular catheter infection. Ann Intern Med 2000 Apr 18;132(8):641-648. [Medline: 10766683]
24. Yilmaz G, Caylan R, Aydin K, Topbas M, Koksal I. Effect of education on the rate of and the understanding of risk factors for intravascular catheter-related infections. Infect Control Hosp Epidemiol 2007 Jun;28(6):689-694. [doi: 10.1086/517976] [Medline: 17520542]
25. Milne JK, Lalonde AB. Patient safety in women's health-care: professional colleges can make a difference. The Society of Obstetricians and Gynaecologists of Canada MORE(OB) program. Best Pract Res Clin Obstet Gynaecol 2007 Aug;21(4):565-579. [doi: 10.1016/j.bpobgyn.2007.01.013] [Medline: 17376746]
26. Gomersall CD, Lam P, Joynt GM. Electronic interactive learning to supplement acute care teaching. Med Educ 2010 May;44(5):526. [doi: 10.1111/j.1365-2923.2010.03694.x] [Medline: 20519008]


## Abbreviations

**AHRQ:** Agency for Healthcare Research and Quality
**BASIC:** Basic Assessment and Support in Intensive Care
**GEE:** Generalized Estimating Equations
**HSOPSC:** Hospital Survey on Patient Safety Culture
**ICU:** Intensive Care Unit
**SBAR:** Situation Background Assessment Recommendation


*Edited by G Eysenbach; submitted 25.11.15; peer-reviewed by WC Su, ML Lypson; comments to author 03.03.16; revised version received 22.04.16; accepted 04.06.16; published 05.07.16.*

*Please cite as:*
*Ling L, Gomersall CD, Samy W, Joynt GM, Leung CCH, Wong WT, Lee A*
*The Effect of a Freely Available Flipped Classroom Course on Health Care Worker Patient Safety Culture: A Prospective Controlled Study*
*J Med Internet Res 2016;18(7):e180*
*URL: http://www.jmir.org/2016/7/e180/*
*doi:10.2196/jmir.5378*
*PMID:27381876*


©Lowell Ling, Charles David Gomersall, Winnie Samy, Gavin Matthew Joynt, Czarina CH Leung, Wai-Tat Wong, Anna Lee. Originally published in the Journal of Medical Internet Research (http://www.jmir.org), 05.07.2016. This is an open-access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/2.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on http://www.jmir.org/, as well as this copyright and license information must be included.



XSL•FO
RenderX

Original Paper

# Possible Biases of Researchers' Attitudes Toward Video Games: Publication Trends Analysis of the Medical Literature (1980–2013)

Aviv Segev[1,2], MD; Mitchell Rovner[3], MD; David Ian Appel[4], MD; Aaron W Abrams[5], MD; Michal Rotem[2], B Med; Yuval Bloch[1,2], MD

[1]Shalvata Mental Health Center, Hod Hasharon, Israel

[2]Sackler School of Medicine, Tel Aviv University, Tel Aviv, Israel

[3]Albert Einstein College of Medicine/Montefiore Medical Center, Bronx, NY, United States

[4]University of Vermont Medical Center, Burlington, VT, United States

[5]SUNY Downstate Medical Center, Brooklyn, NY, United States

**Corresponding Author:**
Aviv Segev, MD
Shalvata Mental Health Center
1st Aliyat Hanoar St
POB 94
Hod Hasharon,
Israel
Phone: 972 9 7478 554
Fax: 972 9 7478 623
Email: segev_a@netvision.net.il

## Abstract

**Background:** The study of video games is expanding, and so is the debate regarding their possible positive and deleterious effects. As controversies continue, several researchers have expressed their concerns about substantial biases existing in the field, which might lead to the creation of a skewed picture, both in the professional and in the lay literature. However, no study has tried to examine this issue quantitatively.

**Objective:** The objective of our study was to examine possible systematic biases in the literature, by analyzing the publication trends of the medical and life sciences literature regarding video games.

**Methods:** We performed a complete and systematic PubMed search up to December 31, 2013. We assessed all 1927 articles deemed relevant for their attitude toward video games according to the focus, hypothesis, and authors' interpretation of the study results, using a 3-category outcome (positive, negative, and neutral). We assessed the prevalence of different attitudes for possible association with year of publication, location of researchers, academic discipline, methodological research, and centrality of the publishing journals.

**Results:** The attitude toward video games presented in publications varied by year of publication, location, academic discipline, and methodological research applied ($P$<.001 for all). Moreover, representation of different attitudes differed according to centrality of the journals, as measured by their impact factor ($P$<.001).

**Conclusions:** The results suggest that context, whether scientific or social, is related to researchers' attitudes toward video games. Readers, both lay and professional, should weigh these contextual variables when interpreting studies' results, in light of the possible bias they carry. The results also support a need for a more balanced, open-minded approach toward video games, as it is likely that this complex phenomenon carries novel opportunities as well as new hazards.

***(J Med Internet Res 2016;18(7):e196)*** doi:10.2196/jmir.5935

**KEYWORDS**

video games; publication trends; bias



XSL•FO
RenderX

## Introduction

Playing video games is a worldwide, significant social phenomenon with possible effects on life and health. Two main attitudes, often polarized, have dominated the interpretation of the consequences of playing video games on well-being since the early days of research in this field [1], with this heated debate continuing today [2]. On one hand, there are those who emphasize the advantages of video games, including beneficial uses of the media [3] such as cognitive enhancement [4,5], rehabilitation [6,7], and prosocial behavior [8]. On the other hand, studies have reported harmful effects of video games on players, including academic deterioration [9], attention and psychosocial problems [10-12], violent behavior [13], and further deleterious effects.

Many possible applications of video games, either commercial or goal-oriented ("serious games"), in the fields of health and medicine were studied in the literature: promoting health behaviors [14,15], motor skills and balance [16,17], cognitive rehabilitation [6,18], medical training [19,20], and even psychotherapy [21,22]. As the literature indicates the possible negative outcomes of video games, and recurrent warnings are being published by leading medical authorities [23], clinicians and policy makers face a complex challenge: to translate the possibilities and applications of video games into clinical practice and official statements, in light of the confusing and contradictory evidence.

These dichotomous views of video games may drive professionals to choose a dichotomous stance, either positive or negative, on video games. These stances have a major impact, especially among physicians, as they may be transmitted, in turn, to their patients and students.

As video game variety, usage, content, and context are widely diverse and have become a part of modern life, Bavelier and Green stated that "One can no more say what the effects of video games are, than one can say what the effects of food are," implying that "the devil is in the details" [2].

Ideally, we look for science to bring forth results and data that will reveal the costs and benefits of this practically universal behavior. However, this optimistic view overlooks the fact that research only answers the hypothesis suggested. Thus, when studying new, emotion-provoking phenomena, the biases of researchers, on which the basic hypotheses are based, might affect and skew the focus of research and the interpretation of its results.

Such biases have been suggested in the literature concerning video games, which might cause a disconnect between the studies' findings and their interpretations in the public and professional literatures [24-26]. Along this line of thought, familiarity with computer games has been identified as a possible moderator of one's beliefs about computer games [27].

In this study, we set forth to examine trends and possible bias in the medical literature focusing on video games, by examining trends by time of publication, country of origin, medical discipline, and research methodology. Revealing such trends may raise awareness of researcher bias, thus helping to formulate a clearer understanding of the interpretation of studies evaluating the risks and rewards of video games. We set forth to examine such biases by examining the researchers' attitudes reflected in the study publication. By attitude, we mean "A settled way of thinking or feeling about something" [28].

## Methods

### Search Strategy

We conducted a systematic search on the PubMed database for all articles published up to December 31, 2013, using synonyms for video games (plural): videogames, video games, "video games"[MeSH] (major and subtypes), electronic games, and "computer games."

### Database Assembly and Variables of Interest

We classified all results according to the following parameters: year of publication, publishing journal, and country of origin (based on the affiliations of the first author). Using the ISI Web of Knowledge (now the Web of Science), we added subject categories to each record, as well as the impact factor of the publishing journal. We then manually accessed each article and read all possible abstracts. If an abstract was not available, or not coherent, we accessed and read the article itself. That process allowed 3 fields to be added to each record: (1) relevance of the article was assessed (relevant/not relevant), based on the role computer games had in the study, because articles may describe video gamers as a control group or as the placebo task for a cognitive test, (2) article type was determined (eg, case study, expert opinion, cross-sectional study, randomized controlled trial), and (3) attitude was assessed on a 3-category variable (as either positive, neutral, or negative), based on the focus, hypothesis, and the article authors' interpretation of the study results and conclusions. Studies hypothesizing that video games increase aggressiveness would be considered to have a negative attitude, but negative results in such a study and a conclusion encouraging doubts about the concept of video game-induced aggressiveness would be considered neutral. On the other hand, a study examining the contribution of active video games to balance rehabilitation would be classified as positive. Similarly, a study examining this issue and reporting a negative result and urging caution when implementing video games in balance rehabilitation would also be classified as neutral.

### Data Quality Assurance

We divided the process of determining the attitude of the article and ensuring interrater reliability into several steps. Initially, 4 classifiers (MR, DIA, AWA, MR) classified 150 articles, each of which was also examined by the lead researcher (AS); all of the researchers discussed conceptual questions. Next in the classification, each of the classifiers worked alone, while every question that arose was discussed with the lead researcher. If considerations about the classification process seemed to have generalization potential, the discussion was relayed to the other classifiers.

After finishing the classification, and to ascertain reliability, a different classifier, who was blinded to the previous categorization, reanalyzed 10% of the sample, randomly selected



XSL•FO
RenderX

by computer. These decisions were compared in order to examine agreement.

### Statistical Analysis

Statistical analysis was performed using IBM SPSS version 16 software (IBM Corp). First, we examined each variable of interest. Spearman correlation examined trends over time. For categorical variables (researchers' location, medical discipline, research methodology, and impact factor), we used chi-square analysis with a follow-up post hoc 2×3 chi-square for each of the subcategories of variables.

## Results

We retrieved 3223 articles in total, of which 170 were duplicates, leaving 3053 articles. For 230 of those (7.53%), we gathered the information from the article itself, as the article had no abstract, or we deemed the abstract not to be informative enough for classification. Only 75 articles (2.46%) had no abstract and the article itself was unattainable, leaving only the title, affiliations, and PubMed's medical subject headings (MeSH) to rely on for classification. In those cases, if the information was not sufficient, we excluded the article from the analysis. For all other articles (n=2748, 90.01%), we based classification on the abstract. We deemed 1126 articles to be irrelevant, as video games were not the focus of the article, leaving 1927 articles, published in 300 different journals, from 64 different countries.

Blinded agreement between researchers on the article's attitude was substantial (κ=.77, $P$<.001). Furthermore, virtually all of the disagreements were either positive versus neutral or negative versus neutral. Only 1 study of the 186 in the verification process was assigned a contrasting attitude (negative-positive); thus, agreement regarding attitude direction was almost perfect (κ=.99, $P$<.001).

Overall, we classified 812 of the 1927 (42.14%) as negative-attitude publications, 301 (15.62%) as neutral, and 814 (42.24) as positive-attitude publications.

### Change in Publication Number and Attitudes Over Time

We found a significant and meaningful Spearman correlation between the number of publications and the year ($r$=.946, $P$<.001), from the first and only publication in 1980, to 312 articles in 2013 (Figure 1).

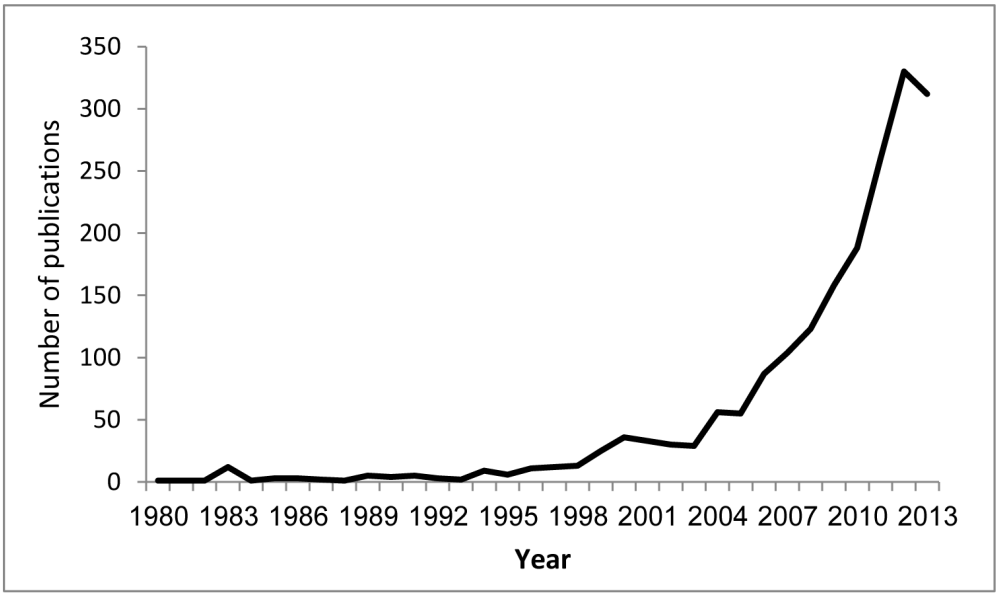
Furthermore, the proportion of video game publications was also positively correlated with the year ($r$=.927, $P$<.001), from 1 out of 279,486 (0.00049%) in 1980, to 312 out of 1,136,703 (0.027%) in 2013, for growth by a factor of 55.1 (Figure 2).

As the number of publications per year was very low until 1999 (<15 per year, for a total of 101 articles in 19 years), we excluded these years from this specific analysis only, as even a single article would create a major shift in the said year. As the years progressed, the proportion of negative publications dropped ($r$=–.907, $P$<.001), while the number of positive and neutral publications increased ($r$=.87, $P$<.001 and $r$=.519, $P$=.047, respectively) (Figure 3).

As we excluded the first 19 years of video game studies due to the scarcity of articles each year, we analyzed those years in 2 large fragments: 1980–1989 and 1990–1998.

Over the first 10 years of research, 34 articles were published, 19 (56%) of them with positive attitude, and 11 (32%) with negative attitude. Over the next 9 years (1990-1998), an additional 67 articles were published, 25 (37%) of them with positive attitude and 32 (48%) with negative attitude.

**Figure 1.** Number of video game-related publications, 1980–2013.






XSL•FO
RenderX

**Figure 2.** Proportion of video game-related publications, 1980–2013.

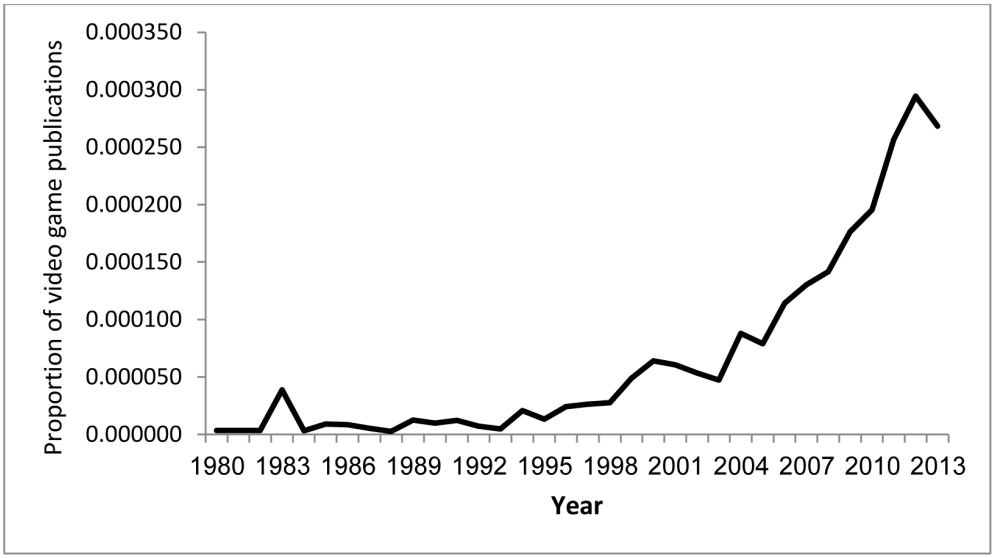
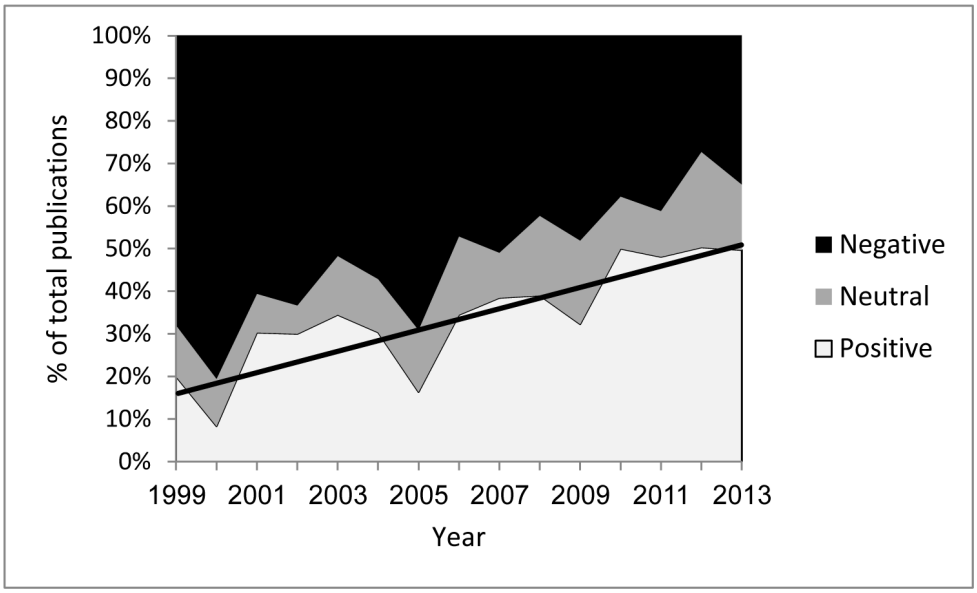



**Figure 3.** Proportions of attitudes toward video games over time, 1999–2013. The line demonstrates the positive linear trend of positive-attitude publications.




## Change in Attitudes Across Regions

As 64 different countries contributed to the pool of articles dealing with video games, we grouped countries according to location (United States, Canada, Europe, Eastern Europe, Australia and New Zealand, Middle East, Asia, Africa, and South America). We combined groups of countries that contributed fewer than 100 articles into 1 group, leaving 6 regions: Australia, Canada, Europe, Asia, United States, and other. A chi-square test revealed a significant association between attitude distribution (positive, negative, or neutral) and the region from which the article originated ($P<.001$). A post hoc test of 2×3 chi-square revealed a significant increase in positive articles (with a reciprocal decrease in negative articles) in the United States ($P<.001$) and Canada ($P=.045$) and a decrease in positive articles (with a reciprocal increase in negative and neutral articles) in Asia ($P<.001$) (Table 1). In Europe there was a trend to increased negative articles and decreased positive articles ($P=.05$).



XSL•FO
RenderX

**Table 1.** Changes in attitudes toward video games across regions, 1980–2013, as reflected in the change in proportion of video game-related articles.

| Region | Countries | Total no. of articles | Change in attitude (%) | | | *P* value |
|---|---|---|---|---|---|---|
| | | | Negative attitude | Neutral attitude | Positive attitude | |
| US | United States | 804 | –5.7% | –0.8% | +6.6 | <.001[a] |
| Canada | Canada | 119 | –9.3% | –1.3% | +10.7% | .045 |
| Australia | Australia, New Zealand | 103 | +4.5% | –3.0% | –1.4% | >.1 |
| Europe | Austria, Belgium, Denmark, Finland, France, Germany, Iceland, Italy, Ireland, Luxembourg, the Netherlands, Norway, Portugal, Scotland, Spain, Sweden, Switzerland, United Kingdom | 554 | +3.9% | +0.3% | –4.1% | .05 |
| Asia | Bangladesh, China, Hong Kong, India, Indonesia, Japan, Korea, Malaysia, Philippines, Singapore, South Korea, Taiwan, Thailand, Vietnam | 176 | +11.9% | +6.0% | –17.8% | <.001[a] |
| Other | Eastern Europe (Bosnia-Herzegovina, Croatia, Czech Republic, Hungary, Lithuania, Poland, Romania, Russia, Serbia, Ukraine); Middle East (Egypt, Greece, Iran, Israel, Lebanon, Oman, Saudi Arabia, Turkey); South America (Argentina, Brazil, Chile, Colombia, Jamaica, Mexico, Puerto Rico); others (Georgia, Nigeria, South Africa, mixed countries) | 171 | +5.9% | –0.4% | –5.4% | >.1 |

[a]Significance remains after Bonferroni correction.

## Change in Attitudes Across Disciplines

Using subject categories, as defined by the Journal Citation Reports (Thomson Reuters, New York, NY), we classified articles by discipline according to the journal in which they were published. Thus, we classified a portion of the 1927 articles in more than 1 domain: 1352 publications (70.16%) were classified to 1 discipline, 490 (25.43%) to 2 disciplines, 38 (1.97%) to 3 disciplines, and 4 (0.21%) to 4 disciplines. We did not assign 43 publications (2.23%) to any discipline, as the publishing journals were not listed in ISI and we could not unequivocally derive the journal discipline from the name of the journal.

The 8 defined disciplines were pediatrics, psychiatry and psychology, neurology, basic sciences, nonmedical and technology, public health and environment, rehabilitation, and internal and general medicine (comprising several fields in medicine in which the number of publications was low, such as ophthalmology, nursing, and family practice). Other than general and internal medicine, the most prominent field was psychiatry and psychology, with 572 (29.68%) of all publications in the field of video games. The next most prominent field was pediatrics (326, 16.92%), followed by public health and environment (311, 16.14%).

A chi-square test revealed a significant association between attitude distribution and disciplines ($P<.001$).

A post hoc test of 2×3 chi-square revealed a reduced number of positive articles (with a reciprocal increase in neutral articles) in psychiatry and psychology ($P<.001$), an increased number of positive articles (with a reciprocal decrease in neutral articles) in general and internal medicine ($P=.001$), an increased number of positive articles (with a reciprocal decrease in neutral and negative articles) both in rehabilitation and in nonmedical and technology domains ($P<.001$ for both), an increased number of negative articles (with a reciprocal decrease in positive articles) in pediatrics ($P<.001$), and an increased number of negative articles (with a reciprocal decrease in neutral articles) in public health and environment ($P<.001$). No correlation was found for neurology and basic sciences (Table 2).



XSL•FO
RenderX

**Table 2.** Change in attitudes toward video games across different disciplines, 1980–2013, as reflected in the change in proportion of video game-related articles.

| Subject | Total no. of articles | Change in attitude (%) | | | *P* value |
|---|---|---|---|---|---|
| | | Negative attitude | Neutral attitude | Positive attitude | |
| Psychiatry and psychology | 572 | +1.1% | +13.9% | –14.9% | <.001[a] |
| Pediatrics | 326 | +17.4% | –1.2% | –16.1% | <.001[a] |
| Rehabilitation | 135 | –39.1% | –10.4% | +49.7% | <.001[a] |
| Nonmedical and technology | 206 | –29.0% | –6.4% | +35.5% | <.001[a] |
| Neurology | 161 | –5.5% | +1.8% | +3.8% | >.1 |
| Basic sciences | 93 | –3.4% | –0.5% | +4.0% | >.1 |
| Public health and environment | 311 | +8.7% | –5.3% | –3.3 % | .001[a] |
| General and internal medicine | 658 | –1.4% | –5.6% | +7.0% | <.001[a] |

[a]Significance remains after Bonferroni correction.

## Change in Attitudes Across Methodological Approaches of Studies

We then divided the articles into 3 categories based on study design: observational studies, interventional studies, and study aggregations (reviews and meta-analyses).

As methodological requirements and evidence-based approaches have shifted greatly since 1980, we examined the correlation between study methodology and the year of publication.

We found a significant and meaningful Spearman correlation between the methodological approach and the year: as the years progressed, the proportion of observational studies declined ($r$=–.75, $P$=.001) and that of interventional studies increased ($r$=.63, $P$=.01). No correlation was found with the number of aggregation-based articles ($r$=–.086, $P$>.1). We then examined whether there was an association between the research methodology and attitude. The correlation was significant ($P$<.001): positive articles increased in interventional studies ($P$<.001), with a reciprocal decline in negative articles, while negative articles increased in observational studies, with a reciprocal decline in positive articles (Table 3).

**Table 3.** Change in attitudes toward video games across methodological approaches, 1980–2013, as reflected in the change in proportion of video game-related articles.

| Type of study | Total no. of articles | Change in attitude (%) | | | *P* value |
|---|---|---|---|---|---|
| | | Negative attitude | Neutral attitude | Positive attitude | |
| Observational | 1081 | +16.0% | +1.1% | –16.9% | <.001[a] |
| Interventional | 686 | –25.8% | –0.9% | +26.8% | <.001[a] |
| Aggregated | 160 | +2.9% | –3.1% | +0.3% | >.1 |

[a]Significance remains after Bonferroni correction.

## Change in Attitude Across Journal Centrality

Based on the impact factor (defined by Journal Citation Reports), we divided the journals into 3 groups: lower impact factor (ranging from 0 to 2.0 or nonlisted), medium impact factor (ranging from 2.001 to 4.0), and high (≥4.001). The groups comprised, respectively, 908 (47.12%), 668 (34.67%), and 351 (18.21%) of the 1927 publications. A chi-square test revealed a significant association between the impact factor group and attitude of the articles ($P$<.001). The low impact factor group tended to publish more positive articles (with a reciprocal decrease in neutral and negative articles) ($P$<.001), whereas both the medium and high impact factor groups tended to publish fewer positive articles, with a reciprocal increase in neutral articles in the medium group ($P$<.001) and a reciprocal increase in negative articles in the high impact factor group ($P$=.008) (Table 4).



XSL•FO
RenderX

**Table 4.** Change in attitudes toward video games across journal centrality, 1980–2013, as reflected in the change in proportion of video game-related articles.

| Impact factor group | Total no. of articles | Change in attitude (%) | | | *P* value |
|---|---|---|---|---|---|
| | | Negative attitude | Neutral attitude | Positive attitude | |
| Low (0–2.0) | 908 | –3.1% | –4.1% | +7.4% | <.001[a] |
| Medium (2.001–4.0) | 668 | +0.4% | +6.6% | –6.9% | <.001[a] |
| High (≥4.001) | 351 | +7.5% | –1.6% | –5.7% | .008[a] |

[a] Significance remains after Bonferroni correction.

# Discussion

## Principal Findings

The number of articles reporting studies of video games is increasing rapidly. It seems that the attitude toward video games is affected by the year of publication, the region of origin of the lead researcher, the discipline from which the article stems, and the research method applied. Moreover, it seems that the representation of different attitudes varies according to the centrality of the journal, as measured by its impact factor.

Surprisingly, in the early years of video games research, while opinion leaders were speaking against video games and their deleterious effects [1], most articles presented a positive attitude. The positive trend prevailing in the 1980s was reversed during the 1990s, when negative attitudes toward video games were reflected in nearly half of the publications. A possible explanation for the proliferation of positive articles in the 1980s is the novelty of this subject, with enthusiastic researchers focusing on this new field and its opportunities. When viewed by year of publication, positive attitudes increased over time (excluding the early "pioneer" years). One possible explanation is that attitudes toward video games are affected by prior experience with the field, and that gaining experience with the medium, and integrating younger researchers who have been exposed to video games all their lives, would lead to a more positive approach. Another possibility would be to view the attitude of the medical research community in a similar way to the patterns of technology adoption. The temporal curves, delineating a rise of positive articles in the 1980s, a strong shift toward negative articles in the 1990s, and then a gradual incline of positive attitudes, follows the trends depicted in Gartner's hype cycles portraying the adoption of new technologies [29].

The assessment based on the country of origin supports the "acquaintance" hypothesis [27]: articles from the United States, a leader in the video game industry, tended to be positive. Articles from Asia, which comprises a mesh of traditional and modern cultures, and with exponential growth of technological penetration and video games, leaned toward negative attitudes. This is possibly a reflection of the repercussions of a rapidly changing culture and assimilation of changes in lifestyle.

Among the different medical disciplines, most of the findings can seem trivial. One could assume that basic science, by its very nature, would tend to lack a polarized attitude. Also, it is not surprising that the field of rehabilitation, which seeks a measure of improvement, would be positively biased. The same bias would be very plausible in the technological disciplines. Public health, though, a discipline that tends to look for risk factors and prevention measures, would understandably be biased toward the more negative attitudes. Neurology seems to be balanced. This leaves 2 disciplines with an intriguing tendency toward the negative: psychiatry and psychology, and, even more so, pediatrics. One can hypothesize that the rapid dissemination of and increase in the number of video games used in the field of pediatrics, along with a more protective and more pronounced generation gap, could be a possible explanation. Another possible explanation would be a specific "toxic" effect specific to early development and mental health. This view is in accord with when video games are being considered as a type of behavioral addiction. However, the positive-attitude studies published in the educational field, as well as in neurology and cognitive rehabilitation, strengthen the suspicion of a negative-attitude bias both in pediatrics and in psychiatry and psychology research.

The methodological partition results may not be surprising. Observational studies are often directed toward negative outcomes (eg, risk factor), while interventional studies usually seek benefits (although difficult to examine systematically, it seems that it is less common for a study to manipulate an intervention that will intentionally cause negative effects than to intervene in order to achieve a more favorable outcome).

The results of this study suggest a possible publication bias as a factor in the basic attitude of the article: negative-attitude articles are more likely to get published in a high impact factor journal. When studying a common phenomenon, such as video games, observational studies (which, as suggested, may favor a negative attitude) offer access to large populations, thus enabling stronger methodology. This might explain the bias of the higher impact factor journals toward publishing studies with a negative attitude. One can expect that, in the coming years, as the technological possibilities of interventional studies improve and as the mass of interventional studies increases, more methodologically robust interventional studies will find their way to more influential journals.

## Limitation

Though the study covered the entire literature indexed in PubMed, our scope was limited to medical and life sciences-related publications. As such, we cannot attest to attitudes in other academic branches dealing with video games, such as education or communications. However, as few such articles have been indexed in PubMed, and consequently analyzed in this study, the general approach toward computer



XSL•FO
RenderX

games seems to be more positive. It should be noted that, as the study focused on attitudes in the medical literature, we searched only the PubMed database. Furthermore, we might have captured additional studies dealing with video games by using different keywords in our search strategy.

An additional limitation that should be noted is that we coded attitudes manually, according to human judgment, which is vulnerable to mistakes and disagreements. However, as the blinded agreement between the authors was very high, it is not likely that misclassification of articles interfered with the results. We chose to focus in this study on video games rather than social media or the internet as a whole, as the topic of video games is grounds for even greater disagreement. As internet use and email can be considered an essential part of the normal modern world (eg, for work, in the household, and for academic assignments), video games are considered "avoidable" and not a necessity, and thus their costs and benefits should be studied more carefully. Although we chose to focus on video games, in reality, the line between social media and video games has become blurred, because, unlike in the past [30], video games now comprise extensive social media and multiplayer options [31], a problem that has been raised in the debate surrounding the new definition of internet gaming disorder in the fifth edition of the *Diagnostic and Statistical Manual of Mental Disorders* [32,33].

## Conclusions

As suggested in the literature [24-26], biases do seem to exist, and recognizing these biases is important for the scientific community studying video games. It allows the reader to put a new study into a wider context, which seems to play a major role, according to our study, and thus could provide a better perspective when interpreting information. Furthermore, these biases should serve as a wake-up call and remind us to keep an open mind about this phenomenon, carrying (as all new phenomena do) both positive and negative perspectives, which are probably intertwined with one another more often than not. Further research should examine possible biases within specific subjects, particularly subjects that are fiercely debated, such as violence, addiction, and physical health implications.

## Acknowledgments

We thank Ms Zmira Silman for her assistance in the statistical analysis and interpretation of results. We received no funding for this study.

## Conflicts of Interest

None declared.

## References

1. Meinel C. Will Pac-Man consume our nation's youth? MIT Technol Rev 1983;May/June(11):28.
2. Bavelier D, Green CS, Han DH, Renshaw PF, Merzenich MM, Gentile DA. Brains on video games. Nat Rev Neurosci 2011 Dec;12(12):763-768 [FREE Full text] [doi: 10.1038/nrn3135] [Medline: 22095065]
3. Griffiths M. Can videogames be good for your health? J Health Psychol 2004 May;9(3):339-344. [doi: 10.1177/1359105304042344] [Medline: 15117533]
4. Dye MW, Green CS, Bavelier D. The development of attention skills in action video game players. Neuropsychologia 2009 Jul;47(8-9):1780-1789 [FREE Full text] [doi: 10.1016/j.neuropsychologia.2009.02.002] [Medline: 19428410]
5. Green CS, Bavelier D. Action-video-game experience alters the spatial resolution of vision. Psychol Sci 2007 Jan;18(1):88-94 [FREE Full text] [doi: 10.1111/j.1467-9280.2007.01853.x] [Medline: 17362383]
6. Franceschini S, Gori S, Ruffino M, Viola S, Molteni M, Facoetti A. Action video games make dyslexic children read better. Curr Biol 2013 Mar 18;23(6):462-466. [doi: 10.1111/j.1467-9280.2007.01853.x] [Medline: 23453956]
7. Zimmermann R, Gschwandtner U, Benz N, Hatz F, Schindler C, Taub E, et al. Cognitive training in Parkinson disease: cognition-specific vs nonspecific computer training. Neurology 2014 Apr 8;82(14):1219-1226. [doi: 10.1212/WNL.0000000000000287] [Medline: 24623840]
8. Greitemeyer T, Osswald S, Brauer M. Playing prosocial video games increases empathy and decreases schadenfreude. Emotion 2010 Dec;10(6):796-802. [doi: 10.1037/a0020194] [Medline: 21171755]
9. Mössle T, Kleimann M, Rehbein F, Pfeiffer C. Media use and school achievement--boys at risk? Br J Dev Psychol 2010 Sep;28(Pt 3):699-725. [Medline: 20849041]
10. Przybylski AK. Electronic gaming and psychosocial adjustment. Pediatrics 2014 Sep;134(3):e716-e722 [FREE Full text] [doi: 10.1542/peds.2013-4021] [Medline: 25092934]
11. Richards R, McGee R, Williams SM, Welch D, Hancox RJ. Adolescent screen time and attachment to parents and peers. Arch Pediatr Adolesc Med 2010 Mar;164(3):258-262. [doi: 10.1001/archpediatrics.2009.280] [Medline: 20194259]
12. Swing EL, Gentile DA, Anderson CA, Walsh DA. Television and video game exposure and the development of attention problems. Pediatrics 2010 Aug;126(2):214-221. [doi: 10.1542/peds.2009-1508] [Medline: 20603258]
13. Strasburger VC, Jordan AB, Donnerstein E. Health effects of media on children and adolescents. Pediatrics 2010 Apr;125(4):756-767. [doi: 10.1542/peds.2009-2563] [Medline: 20194281]



XSL•FO
RenderX

14. Read JL, Shortell SM. Interactive games to promote behavior change in prevention and treatment. JAMA 2011 Apr 27;305(16):1704-1705. [doi: 10.1001/jama.2011.408] [Medline: 21447802]
15. Thompson D, Bhatt R, Lazarus M, Cullen K, Baranowski J, Baranowski T. A serious video game to increase fruit and vegetable consumption among elementary aged youth (Squire's Quest! II): Rationale, design, and methods. JMIR Res Protoc 2012;1(2):e19 [FREE Full text] [doi: 10.2196/resprot.2348] [Medline: 23612366]
16. Sabel M, Sjölund A, Broeren J, Arvidsson D, Saury J, Blomgren K, et al. Active video gaming improves body coordination in survivors of childhood brain tumours. Disabil Rehabil 2016 Jan 5:1-12. [doi: 10.3109/09638288.2015.1116619] [Medline: 26728453]
17. Imam B, Miller WC, Finlayson H, Eng JJ, Jarus T. A randomized controlled trial to evaluate the feasibility of the Wii Fit for improving walking in older adults with lower limb amputation. Clin Rehabil 2015 Dec 30. [doi: 10.1177/0269215515623601] [Medline: 26721873]
18. Sue D, Ray P, Talaei-Khoei A, Jonnagaddala J, Vichitvanichphong S. Assessing video games to improve driving skills: a literature review and observational study. JMIR Serious Games 2014;2(2):e5 [FREE Full text] [doi: 10.2196/games.3274] [Medline: 25654355]
19. de Araujo Thiago Bozzi TB, Silveira FR, Souza DL, Strey YT, Flores CD, Webster RS. Impact of video game genre on surgical skills development: a feasibility study. J Surg Res 2016 Mar;201(1):235-243. [doi: 10.1016/j.jss.2015.07.035] [Medline: 26850208]
20. Nicolaidou I, Antoniades A, Constantinou R, Marangos C, Kyriacou E, Bamidis P, et al. A virtual emergency telemedicine serious game in medical training: a quantitative, professional feedback-informed evaluation study. J Med Internet Res 2015;17(6):e150 [FREE Full text] [doi: 10.2196/jmir.3667] [Medline: 26084866]
21. Ceranoglu TA. Star Wars in psychotherapy: video games in the office. Acad Psychiatry 2010;34(3):233-236. [doi: 10.1176/appi.ap.34.3.233] [Medline: 20431107]
22. Horne-Moyer HL, Moyer BH, Messer DC, Messer ES. The use of electronic games in therapy: a review with clinical implications. Curr Psychiatry Rep 2014 Dec;16(12):520 [FREE Full text] [doi: 10.1007/s11920-014-0520-6] [Medline: 25312026]
23. Council OC. From the American Academy of Pediatrics: policy statement--media violence. Pediatrics 2009 Nov;124(5):1495-1503. [doi: 10.1542/peds.2009-2146] [Medline: 19841118]
24. [No authors listed]. Is exposure to media violence a public-health risk? Lancet 2008 Apr 5;371(9619):1137. [doi: 10.1016/S0140-6736(08)60495-X] [Medline: 18395556]
25. Ferguson CJ. Violent video games and the Supreme Court: lessons for the scientific community in the wake of Brown v. Entertainment Merchants Association. Am Psychol 2013;68(2):57-74. [doi: 10.1037/a0030597] [Medline: 23421606]
26. Ferguson CJ, Kilburn J. Much ado about nothing: the misestimation and overinterpretation of violent video game effects in eastern and western nations: comment on Anderson et al. (2010). Psychol Bull 2010 Mar;136(2):174-8; discussion 182. [doi: 10.1037/a0018566] [Medline: 20192554]
27. Przybylski AK. Who believes electronic games cause real world aggression? Cyberpsychol Behav Soc Netw 2014 Apr;17(4):228-234. [doi: 10.1089/cyber.2013.0245] [Medline: 24256132]
28. Oxford Dictionaries: Attitude. Oxford, UK: Oxford University Press; 2016. URL: http://www.oxforddictionaries.com/definition/english/attitude?q=Attitude [accessed 2016-06-12] [WebCite Cache ID 6iDD0CszR]
29. Linden A, Fenn J. Understanding Gartner's Hype Cycles: Strategic Analysis Report No. R-20-1971. Stamford, CT: Gartner, Inc; 2003 May 30. URL: http://www.ask-force.org/web/Discourse/Linden-HypeCycle-2003.pdf [accessed 2016-07-11] [WebCite Cache ID 6ivK2sKlQ]
30. Braun C, Giroux J. Arcade video games: proxemic, cognitive and content analyses. J Leisure Res 1989;21(2):92-105.
31. Lee HJ, Choi J, Kim JW, Park SJ, Gloor P. Communication, opponents, and clan performance in online games: a social network approach. Cyberpsychol Behav Soc Netw 2013 Dec;16(12):878-883 [FREE Full text] [doi: 10.1089/cyber.2011.0522] [Medline: 23745617]
32. King DL, Delfabbro PH. Issues for DSM-5: video-gaming disorder? Aust N Z J Psychiatry 2013 Jan;47(1):20-22. [doi: 10.1177/0004867412464065] [Medline: 23293310]
33. Starcevic V. Is Internet addiction a useful concept? Aust N Z J Psychiatry 2013 Jan;47(1):16-19. [doi: 10.1177/0004867412461693] [Medline: 23293309]


## Abbreviations

**MeSH:** Medical Subject Headings



XSL•FO
RenderX

*Edited by G Eysenbach; submitted 04.05.16; peer-reviewed by R Buday, J Apolinário-Hagen; comments to author 26.05.16; revised version received 12.06.16; accepted 17.06.16; published 18.07.16.*

*Please cite as:*
*Segev A, Rovner M, Appel DI, Abrams AW, Rotem M, Bloch Y*
*Possible Biases of Researchers' Attitudes Toward Video Games: Publication Trends Analysis of the Medical Literature (1980–2013)*
*J Med Internet Res 2016;18(7):e196*
*URL: http://www.jmir.org/2016/7/e196/*
*doi:10.2196/jmir.5935*
*PMID:27430187*

©Aviv Segev, Mitchell Rovner, David Ian Appel, Aaron W Abrams, Michal Rotem, Yuval Bloch. Originally published in the Journal of Medical Internet Research (http://www.jmir.org), 18.07.2016. This is an open-access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/2.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on http://www.jmir.org/, as well as this copyright and license information must be included.



XSL•FO
RenderX

Original Paper

# Older Cancer Patients' User Experiences With Web-Based Health Information Tools: A Think-Aloud Study

Sifra Bolle[1], MSc; Geke Romijn[1,2], MSc; Ellen M A Smets[3], PhD; Eugene F Loos[1], PhD; Marleen Kunneman[3,4], PhD; Julia C M van Weert[1], PhD

[1]Amsterdam School of Communication Research/ ASCoR, Department of Communication Science, University of Amsterdam, Amsterdam, Netherlands
[2]Faculty of Behavioural and Movement Sciences, Department of Clinical Psychology, VU University, Amsterdam, Netherlands
[3]Department of Medical Psychology, Amsterdam Medical Center/ AMC, University of Amsterdam, Amsterdam, Netherlands
[4]Department of Medical Decision Making, Leiden University Medical Center, Leiden, Netherlands

**Corresponding Author:**
Sifra Bolle, MSc
Amsterdam School of Communication Research/ ASCoR
Department of Communication Science
University of Amsterdam
P.O. Box 15791
Amsterdam, 1001 NG
Netherlands
Phone: 31 205254886
Fax: 31 205253681
Email: S.Bolle@uva.nl

**Related Articles:**

Comment in: http://www.jmir.org/2016/11/e273/

Comment in: http://www.jmir.org/2016/11/e289/

## Abstract

**Background:** Health information is increasingly presented on the Internet. Several Web design guidelines for older Web users have been proposed; however, these guidelines are often not applied in website development. Furthermore, although we know *that* older individuals use the Internet to search for health information, we lack knowledge on *how* they use and evaluate Web-based health information.

**Objective:** This study evaluates user experiences with existing Web-based health information tools among older (≥ 65 years) cancer patients and survivors and their partners. The aim was to gain insight into usability issues and the perceived usefulness of cancer-related Web-based health information tools.

**Methods:** We conducted video-recorded think-aloud observations for 7 Web-based health information tools, specifically 3 websites providing cancer-related information, 3 Web-based question prompt lists (QPLs), and 1 values clarification tool, with colorectal cancer patients or survivors (n=15) and their partners (n=8) (median age: 73; interquartile range 70-79). Participants were asked to think aloud while performing search, evaluation, and application tasks using the Web-based health information tools.

**Results:** Overall, participants perceived Web-based health information tools as highly useful and indicated a willingness to use such tools. However, they experienced problems in terms of usability and perceived usefulness due to difficulties in using navigational elements, shortcomings in the layout, a lack of instructions on how to use the tools, difficulties with comprehensibility, and a large amount of variety in terms of the preferred amount of information. Although participants frequently commented that it was easy for them to find requested information, we observed that the large majority of the participants were not able to find it.

**Conclusions:** Overall, older cancer patients appreciate and are able to use cancer information websites. However, this study shows the importance of maintaining awareness of age-related problems such as cognitive and functional decline and navigation



XSL•FO
RenderX

difficulties with this target group in mind. The results of this study can be used to design usable and useful Web-based health information tools for older (cancer) patients.

***(J Med Internet Res 2016;18(7):e208)*** doi:10.2196/jmir.5618

**KEYWORDS**

user experience; eHealth; usability; think aloud; aging; cancer

## Introduction

An increasing amount of health information is delivered on the Internet [1]. At the same time, more and more patients search the Internet to find information regarding their illness or treatment [2]. This is a fortunate development as the use of Web-based health information tools (eg, Web-based patient education, patient portals, and health-related apps) improves patients' health-related outcomes [3]. These tools can serve different functions, such as providing information or improving communication with health care providers through the use of so-called "preparatory tools" that support the patient in preparing for consultations and/or in making treatment decisions. Examples of preparatory tools are question prompt lists (QPLs) and decision aids. QPLs are structured lists of questions or topics that patients can use to prepare for a medical encounter by choosing questions they would like to ask their provider during the consultation. QPLs have been found to enhance patient participation and improve emotional and cognitive outcomes in cancer patients [4]. Decision aids are tools that help patients make decisions about their treatment by informing them of treatment options and helping them clarify their values. This helps patients communicate their values and wishes to their health care provider who can use this information to create an optimal treatment plan tailored to the patient [5].

Older patients are an important target group for Web-based health information considering the fact that many diseases (eg, cancer, diabetes, and hypertension) are diseases of older adults [6]. A recent literature review revealed that an increasing number of Web-based health information tools for older patients have been developed and that older patients benefit from the use of these tools as evidenced by improved outcomes such as self-efficacy, blood pressure, hemoglobin levels, and cholesterol levels [7]. These results were especially prevalent for Web-based health information tools with a variety of functions. However, descriptions of development processes are often not published, raising questions about the extent to which these Web-based health information tools are optimally adapted to older patients' needs and abilities [8]. Hence, we know *that* older individuals use the Internet to search for health information and that this may result in positive health outcomes, but we lack knowledge on *how* they use and evaluate Web-based health information. Older patients experience more difficulties using Web-based technologies compared with younger age groups as they are simply less experienced in using Web-based technologies. Although this problem might resolve itself in the future decades as new generations of older adults have more experience with Web-based technologies, it is to be expected that future older generations will still face usability issues due to age-related problems such as cognitive decline and sensory and functional limitations [9,10]. First, sensory limitations such as visual decline can affect the readability of a website, for example, when small font sizes are used. Second, functional limitations such as the worsening of fine motor skills can cause problems when precise mouse movements are required, for example, using pull-down menus, which only stay open when someone moves over the area with the mouse [11]. Therefore, the use of static navigational elements, such as drop-down menus that stay open until one clicks on a link, has been recommended [9,12]. Third, (age related) cognitive decline can hinder someone's ability to process information. The more information that is presented on websites, the more difficult it becomes for people with cognitive decline to find required information [12]. For example, Czaja et al [13] demonstrated that the influence of age on the use of technology (ie, computer and Internet use) is mediated by such age-related problems. Involvement of the end user at an early stage in the development process for Web-based health information tools is of high importance to tailor the tools to address such problems [14,15]. Still, literature on user experiences with Web-based health information tools for patients is scarce [16,17]. As a result, cumulative knowledge to be used for the development of Web-based health information tools for older patients is largely missing.

Many studies consider usability, that is, "the extent to which a product can be used by specified users to achieve specified goals with effectiveness, efficiency, and satisfaction," as the main outcome to evaluate Web-based health information tools [18]. However, usability is only 1 dimension of the user experience [19]. According to the technology acceptance model, technology acceptance and usage can be predicted by ease of use (ie, usability) and perceived usefulness [20,21]. Putting user experience in the context of Web-based health information tools, for which a patient is the end user, we therefore argue that we must evaluate not only *usability* but also *perceived usefulness* in terms of content and intention to use the tool. Where usability is related to the ease of use of a system, perceived usefulness addresses "the degree to which a person believes that using a particular system would enhance his or her job performance" [22]. The aim of this study is, therefore, to evaluate user experience (ie, *usability* and *perceived usefulness*) of Web-based health information tools among older patients. Important aspects of *usability* are the extent to which the tool meets the patients' needs and abilities in terms of *navigation strategy* and *navigation problems* [23]. A Web-based health information tool high in *perceived usefulness* delivers its *content* in a way that satisfies the information needs of the user and increases their *intention* to use [24]. First, the *content* of Web-based health information tools should be considered in user experience evaluations for Web-based health information tools for older patients as the information needs of older patients might differ from those of younger patients [25]. Second, older patients who



XSL•FO
RenderX

might be used to receiving health information through traditional media sources must perceive the Web-based information to be useful to develop an *intention* to actually use it.

As cancer is frequently a disease among older people [6], we will assess user experience with existing Web-based health information tools among older (≥ 65 years) cancer patients and survivors and their partners. We selected 7 Web-based health information tools with different functions (ie, information provision tools and preparatory tools). The results of this study can be used in the systematic development of Web-based health information tools for older cancer patients.

## Methods

### Study Design, Setting, and Sample

This study is part of a larger project in which we systematically developed a Web-based health information tool for older colorectal cancer patients. Participants were recruited from PanelCom, a panel of cancer patients who participated in previous studies with our research group and consented to be contacted again to participate in future studies. Participants were included if they were: (1) aged 65 years or older and (2) had been diagnosed with colorectal cancer or were a partner of a colorectal cancer patient or survivor. Approval for the study was obtained from the Institutional Review Board of the University of Amsterdam (2014-CW-64).

Think-aloud observations are a classic method to assess user experience of Web-based interfaces [26]. As older individuals might have short-term memory problems, valuable data might get lost when asking participants questions after using Web-based health information tools. The think-aloud methodology allows us to observe the actual reactions of the participants during the use of the tools. Another advantage is that the think-aloud method requires low numbers of participants. Throughout the literature, it has been found that only 5 to 9 participants can detect 80% to 90% of usability problems on a website [27-29]. However, the think-aloud method has also been criticized with respect to the validity of the self-reported data that it generates [30]. Previous research has therefore suggested combining think-aloud data with observational data [31]. Therefore, we recorded all sessions by video to be able to systematically observe how participants used the websites and preparatory tools. The think-aloud method enabled us to identify usability problems via observation and self-report. Moreover, the interview setting enabled us to query the participant concerning the perceived usefulness of the tools, specifically with regard to the content and intention to use the tool (see "Materials"). This combination of think-aloud data and interview data has been used previously to investigate usability and perceived usefulness [32].

### Materials

#### *Cancer Information Websites*

To identify characteristics of cancer information websites that best match the needs of the target group, we selected 3 existing websites that cancer patients might find when searching for information on the Internet, but the sources offering the information differed. When searching for Web-based health information, people commonly use general search engines such as Google, use short phrases or keywords, and tend not to look further than the first page of the search results [33]. We therefore selected a website that is the first hit on Google in the Netherlands when searching for the Dutch word for chemotherapy, which is 1 of the 3 most used treatments for cancer in the Netherlands [34]. This is a website with general information on chemotherapy that is owned by a pharmacist (website 1; [35]). As many Web-based health information consumers have difficulties in assessing the credibility of Web-based information [33], we next searched for a website from a seemingly reliable source, specifically a hospital. When searching for the Dutch words for "cancer" and "hospital," the first hit on Google refers to a website for a specialized hospital for cancer patients in the Netherlands (The Antoni van Leeuwenhoek Hospital; website 2; [36]). As this study is part of a larger project in which we systematically developed a Web-based health information tool for older colorectal cancer patients, we selected the website of an expertise center for gastrointestinal cancer in the Netherlands (Gastrointestinal Oncology Cancer Center Amsterdam; website 3; [37]). Furthermore, we made sure in selecting the 3 websites that they differed from each other in terms of offering different modalities (ie, textual, visual, and audiovisual information) through which the information was presented and that they differed from each other with respect to various usability recommendations (eg, minimum 12-point font size, a button to increase text size, and static navigational elements), as proposed by Pernice and Nielsen [9]. Website 1 provided patients with textual information and used illustrations that clarified the text. The text on this website had a font size larger than 12 points but did not have the option to increase text size. The website did not have static navigational elements. Website 2 contained textual and audiovisual information. The text of this website had a smaller font size than 12 points and did not have the option to increase text size. The website did have static navigational elements. Website 3 contained textual and audiovisual information. The text on this website had a font size smaller than 12 points but had a button to increase text size. This website also had static navigational elements (eg, links and menus that do not change or move).

#### *Question Prompt Lists*

We used 3 Dutch Web-based QPLs for cancer patients. The first QPL was integrated into the website with chemotherapy information described previously (QPL 1; [38]). On this website, 4 QPLs were available concerning "preparation," "a good conversation," "side effects," and "after the treatment." The QPLs were in PDF format and were no longer than 1 page. The QPLs consisted of questions that patients might ask during consultations. Questions could be selected by ticking a checkbox in front of each question.

The second QPL was developed by the Dutch Breast Cancer Association for breast cancer patients and their family members (QPL 2; [39]). The homepage of the QPL contained 11 buttons that consisted of the main themes of the QPL and 3 other buttons for explanations and instructions, advice on preparing for consultations, and contact with an expert. The 11 main themes



XSL•FO
RenderX

were further divided into 86 subthemes. The main themes were "diagnosis and treatment," "questions for family members," "hereditary and familial breast and ovarian cancer," "breast cancer among older patients," "breast cancer among younger patients," "symptoms of breast cancer," "work and re-integration," "breast reconstruction," "metastasized breast cancer," "nutrition and exercise," and "breast cancer among men."

The third QPL was developed by researchers from the Academic Medical Center in Amsterdam for patients with esophageal cancer to prepare for their first consultation with the surgeon after surgery (as this tool was developed for research purposes and is not publicly available, we included a screenshot in Multimedia Appendix 1: QPL3). This QPL started with an explanation of the goal and content of the QPL and gave instructions to use the QPL. The QPL contained 76 questions across 9 themes: "operation and hospitalization," "additional care," "physical activity," "social or emotional problems," "nutrition," "the probe," "the future," "physical assumptions," and "medical care." In addition, users could add their own questions.

### *Decision Aid: Values Clarification Tool*

To the best of our knowledge, there was no publicly available Web-based decision support tool for cancer patients available at the start of our study. Therefore, we used a decision aid developed by researchers at the Leiden University Medical Center (this tool has been previously used for study purposes only; see Multimedia Appendix 2). This decision aid uses the values clarifications method, which aims to encourage the consideration of all relevant treatment options and/or attributes of options while lowering the processing burden so patients can adequately identify and integrate their values in forming a preference [40]. Values clarification methods can aid older patients to individually tailor treatment decision making according to their life values. The values clarification tool aimed to assess the relative importance of rectal cancer treatment outcomes. Patients were first asked to rate the importance of the occurrence of the best and worst probability of each possible treatment outcome (all else being equal) on a 4-point scale ranging from "not at all important" to "very important." Next, patients were asked to rate the importance of 10 paired outcome scenarios on a 7-point scale, ranging from "a strong preference for scenario 1" to "a strong preference for scenario 2." An example of a paired scenario was "Scenario 1: Fecal incontinence. Out of 100 people: 65 will have this, 35 will not. Sexual problems. Out of 100 people: 60 will have this, 40 will not. Scenario 2: Fecal incontinence. Out of 100 people: 50 will have this, 50 will not. Sexual problems. Out of 100 people: 70 will have this, 30 will not." The questions in the values clarification tool were adaptive conjoint analysis based, meaning that the paired scenarios were tailored to each individual patient based on what they consider important tradeoffs [5].

## Procedure

Each participant evaluated the usability and perceived usefulness of 3 Web-based health information tools (ie, 1 of the 3 websites providing information, 1 of the 3 QPLs, and the values clarification tool; see Materials). Participants were first allocated to 1 of the 3 cancer information websites. We strove for an equal distribution of gender and being a cancer patient or survivor or a partner. All participants used the tools individually. Participants who were assigned to website 1 (ie, the website providing information on chemotherapy treatment for cancer) were also assigned to QPL 1, as this QPL was part of the same cancer information website. As QPL 2 was designed for female breast cancer patients, only female participants were assigned to this tool. As we had only 1 values clarification tool, all participants evaluated their user experience with this tool (see Table 1 for the distribution of participants across the tools).

We visited the participants at their homes. The sessions started with an explanation of the procedure, signing informed consent, and a short survey that assessed demographic information (ie, age, gender, and education), illness-related information (ie, diagnosis and treatment), and computer experience (ie, amount and purpose of computer use and usage of Web-based health information tools). Education was divided into low level of education (ie, primary education, lower vocational education, preparatory secondary vocational education, and intermediate secondary vocational education), middle level of education (ie, senior secondary vocational education and university preparatory vocational education), and high level of education (ie, higher vocational education and university). We provided all participants with the same hardware, using the same settings. Participants were instructed to perform several tasks according to the protocol. Participants were explicitly instructed to think aloud while executing tasks. It was emphasized that the goal of these tasks was not to test the quality of their computer skills but rather to test the usability of the Web-based health information tools. After finishing the tasks in the protocol, participants received a monetary reward of €20 for their participation.

To assess user experience, we developed an interview protocol containing different tasks (ie, search tasks, application tasks and evaluations; see Textbox 1). Some search tasks aimed to obtain insight in terms of the general navigation behavior of participants and contained the instruction to imagine a certain scenario and search for information one would like to receive in a particular situation. Other search tasks contained more elaborate instructions to search for specific information to assess information preferences. Evaluation tasks offered participants the opportunity to give their opinion about the content and usefulness for (parts of) the website or tool. Application tasks provided information about how participants use the website or preparatory tool. The protocol changed depending on the content of the Web-based health information tool visited for 1 search task (see Textbox 1; search task 7). The amount and types of tasks and questions remained the same.



XSL•FO
RenderX

**Textbox 1.** Description of questions in the observation protocol.

1. Open the website (application task)

2. What is your first impression of this website? (evaluation)

3. Imagine you just got diagnosed with cancer. What information would you like to find on this website? Try to find that information (search task)

4. Were you able to find the information? Was it easy to find the information? What made it easy or difficult? Was the information understandable? (evaluation)

5. Go to the homepage (application task)

6. Was it easy to go back to the homepage? (evaluation)

7. Try to find (search task):

- information on how to prepare for a consultation with your health care provider (website 1)

- information on colorectal cancer (website 2)

- experiences of other patients on this website (website 3)

8. Were you able to find the information? Was it easy to find the information?

9. What made it easy or difficult? Was the information understandable? (evaluation)

10. Would you use this tool in the case that you were a patient for whom this tool is designed? (evaluation—intention to use)

## Analysis

All think-aloud observations were transcribed and coded independently by 2 researchers. Disagreements were resolved through discussion. We analyzed user experience on the basis of 2 dimensions: usability and perceived usefulness. Regarding the *usability*, the data from the think-aloud protocols were analyzed from 2 different perspectives as suggested by Van Waes [23]: (1) *navigation strategy* (ie, which navigation tools did the participant use?) and (2) *navigation problems* (ie, what were the navigation barriers the participant came across?). During the first round of coding, we initially used these 2 perspectives as coding categories.

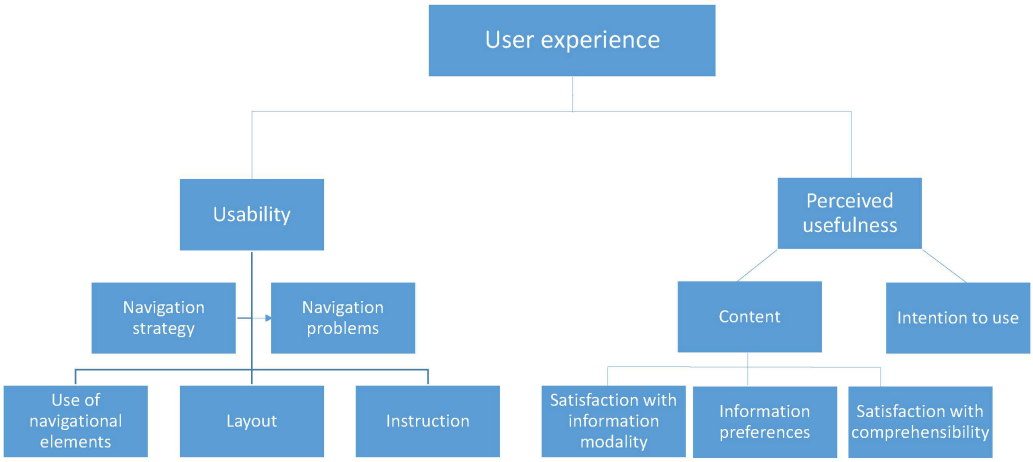
All comments regarding usability could be classified under these codes, but as navigation strategy often led to navigation problems, we combined the 2 codes into 1 code: navigation strategy and problems. We subsequently identified 3 categories of navigation strategies and problems: (1) use of navigational elements, (2) layout, and (3) instructions. These 3 categories were used as subcodes during the second round of coding; all comments regarding usability could be classified under these subcodes.

Regarding the *perceived usefulness* of the Web-based health information tools, we coded whether participants had negative or positive remarks regarding the *content* presented on the website and whether participants had an *intention* to use the tools.

Regarding the negative and positive remarks regarding the content presented on the website, we identified 3 subcodes during the first round of coding: (1) satisfaction with information modality, (2) information preferences, and (3) satisfaction with comprehensibility. During the second round of coding, all positive and negative remarks regarding the content could be classified under these subcodes.

Regarding intention to use the tools, we coded whether and why participants indicated that they would or would not use the tool in the case they were a cancer patient or a partner of a cancer patient again. The codetree is shown in Figure 1.

**Figure 1.** Code tree.






XSL•FO
RenderX

## Results

### Participants and Their Characteristics

Participants were (colorectal) cancer survivors (diagnosed more than 2 years ago; n=12), colorectal cancer patients (diagnosed less than 2 years ago; n=3), and their partners (n=8). The median age of the participants was 73 (interquartile range 70-79). Participants' characteristics are listed in Table 1.

**Table 1.** Participants' characteristics.

| | W1[a] (n=6) n (%) | W2[b] (n=8) | W3[c] (n=9) | QPL 1[d] (n=6) | QPL 2[e] (n=7) | QPL 3[f] (n=10) | Total or values clarification[g] (n=23) |
|---|---|---|---|---|---|---|---|
| | n (%) | n (%) | n (%) | n (%) | n (%) | n (%) | n (%) |
| **Gender** | | | | | | | |
| Male | 4 (75) | 2 (25) | 5 (67) | 4 (67) | 0 (0) | 7 (70) | 11 (48) |
| Female | 2 (25) | 6 (75) | 4 (33) | 2 (33) | 7 (100) | 3 (30) | 12 (52) |
| **Age** | | | | | | | |
| Median | 74.5 | 76 | 70 | 74.5 | 73 | 73.5 | 73 |
| IQR[h] | 73-77.5 | 72.25-81.25 | 67.5-79.5 | 73-77.5 | 72-82 | 67.5-79 | 70-79 |
| **Education** | | | | | | | |
| Low | 2 (33) | 3 (38) | 4 (44) | 2 (33) | 3 (43) | 4 (40) | 9 (39) |
| Middle | 1 (17) | 2 (25) | 1 (11) | 1 (17) | 2 (29) | 1 (10) | 4 (17) |
| High | 3 (50) | 2 (25) | 4 (44) | 3 (50) | 2 (29) | 4 (40) | 9 (39) |
| Other | | 1 (13) | | | | 1 (10) | 1 (4) |
| **Diagnosis** | | | | | | | |
| Colorectal cancer (patient) | 2 (33) | 1 (13) | 0 (0) | 2 (33) | 0 (0) | 1 (10) | 3 (13) |
| Colorectal cancer (survivor) | 2 (33) | 2 (25) | 6 (67) | 2 (33) | 2 (29) | 6 (60) | 10 (44) |
| Other cancer (patient) | 0 (0) | 0 (0) | 1 (11) | 0 (0) | 0 (0) | 2 (20) | 2 (9) |
| Other cancer (survivor) | 1 (17) | 1 (13) | 0 (0) | 1 (17) | 0 (0) | 0 (0) | 1 (4) |
| No diagnosis (ie, partners) | 1 (17) | 4 (50) | 2 (22) | 1 (17) | 5 (71) | 1 (10) | 7 (30) |
| **Computer use per week[i] (in hours)** | | | | | | | |
| 0-2 | 3 (50) | 6 (67) | 5 (63) | 3 (50) | 5 (71) | 6 (60) | 14 (61) |
| 2-10 | 1 (17) | 1 (11) | 0 (0) | 1 (17) | 0 (0) | 1 (10) | 3 (13) |
| 10+ | 2 (33) | 2 (22) | 3 (38) | 2 (33) | 2 (29) | 3 (30) | 6 (26) |

[a] Website 1: website about chemotherapy.

[b] Website 2: website for Antoni van Leeuwenhoek Hospital.

[c] Website 3: website for Gastrointestinal Oncology Center Amsterdam.

[d] QPL 1: QPL on www.chemotherapie.nl.

[e] QPL 2: QPL of the Dutch Breast Cancer Association.

[f] QPL 3: developed by researchers of the Academic Medical Center in Amsterdam for patients with esophageal cancer.

[g] Values clarification tool: a values clarification tool developed by researchers at the Leiden University Medical Centre.

[h]IQR: interquartile range.

[i] Computer use: personal computer and/or tablet.



XSL•FO
RenderX

## Usability

### *Navigation Strategy and Problems*

Although participants frequently commented that it was easy for them to find requested information (n=16; 70%), we observed that the large majority of the participants encountered problems in their navigation strategy and hence were not able to find the requested information (n=21; 91%). We identified 3 categories of navigation strategies that led to problems in optimally navigating the Web-based health information tools: (1) use of navigational elements, (2) layout, and (3) instructions.

### *Use of Navigational Elements*

Participants often started to search for information in the center of a webpage (n=18; 78%) without paying attention to the structure of the website (ie, using a menu on the website to search for information). Website 3, for example, presented a large amount of information in the center of the webpage. The text contained several clickable links to other webpages. When we asked participants to search for specific information, most participants read the text in the center of the webpage and clicked on links provided in the text. They did not consider the menus at the top and on the left side of the webpage (n=8; 89%). Some participants commented that websites with 2 or more menu bars were too complex (n=5; 22%). Only 1 participant (4%) wanted to use the search bar to search for information but could not find the search bar.

For all Web-based health information tools, participants were required to scroll down to see an entire webpage. Overall, participants were able to and did not mind scrolling up and down (n=19; 83%), although 2 participants (9%) commented that the structure of the homepage would be easier to understand if they did not need to scroll.

QPL 2 presented a pop-up after the first question was selected. The pop-up presented the option to save the questions and to send the selected questions by email. However, this pop-up was confusing for some participants, as they had the feeling that the pop-up was an error message and that they did something wrong (n=2; 29%). Although there was an option to "continue anonymously," some participants did not understand how to return to the QPL without saving the questions or leaving their email address (n=3; 43%).

When participants were given the task of returning to the homepage, they mostly used the arrow at the left top corner of the browser (n=12; 61%). Participants were not aware of the possibility of returning to the homepage of the website by clicking on the home button or on the logo of the website (n=17; 74%). Only website 1 had a "home" button to return to the homepage. However, this button was very small, and only 1 participant noticed it. One participant commented that they wanted to have a button with the text "back to the previous page" or "back to the homepage." Website 3 had a button titled "back to care." This button did not lead back to the previous page but to a completely different page on the website, which was confusing (n=5; 56%).

QPL 3 had 2 navigation possibilities. The first possibility was to go through all the questions in the QPL consecutively. The second option was to click on themes that were of interest to the participant and select the questions that were categorized under the specific theme and proceed by clicking on the next theme that was of interest. Participants mostly used the first option (n=9; 90%). Although participants went through the different themes and questions one by one, 2 (20%) did mention that they appreciated the subdivision into themes. The values clarification tool presented participants with one question at a time, which did not cause navigation difficulties.

### *Layout*

Some participants were not able to read the text due to small font sizes (n=6; 26%) and/or a lack of contrast (n=5; 22%). Participants were not aware of the option to increase font size or were not able to find this option (n=2; 25%) that was presented by website 2. When we gave these participants instructions on how to increase the font size, they did appreciate this function.

Website 2 had a background consisting of a blurred illustration. Two participants commented that this was distracting because they did not know whether the illustration was blurred on purpose or whether this was due to their own visual decline (25%). The other 2 websites had plain backgrounds with colors that contrasted the text, which was greatly appreciated by the participants. Website 1 used different shades of brown colors. One participant (17%) mentioned that it was too difficult to see the differences, which made it difficult to read the text and to distinguish between buttons.

Website 1 and the values clarification tool presented buttons that were too small and too close together, resulting in participants clicking on the wrong button (n=7; 24%). Sometimes, participants were not aware of clicking the wrong button, leading to confusion as they saw a webpage with information that they did not expect or could not continue using the values clarification tool (n=5; 22%).

QPLs 1 and 3 used checkboxes that could be clicked on to select a question. QPL 2 used "+" and "-" symbols to select or deselect a question. These symbols were not always clear for participants (n=3; 43%). In addition, the same participants did not see that the question was added to their checklist after they clicked on it and when the "+" symbol changed into a "–" symbol. Furthermore, in the same QPL, the selected question changed from a black font into a gray font. Some participants did not notice this or were unable to see this change in colors (n=3; 43%).

The values clarification tool had a colored progress bar, which was appreciated by 2 participants (9%) but not noticed by the rest of the participants. One participant (4%) was color blind and could not see the progress in the bar.

### *Instructions*

QPL 2 presented users with short instructions to help them with navigation while using the tool. Participants appreciated these instructions (n=2; 29%). One participant (14%) commented that they wanted instructions to navigate the website, for example, an instruction such as "click here if you want to have information on this topic." QPL 3 started with an instruction



XSL•FO
RenderX

on how to use and navigate through the tool. However, given that the instructions disappeared when participants were using the tool, some participants forgot these instructions (n=3; 30%). The instruction text was also considered too long, which resulted in some participants lacking the motivation to read the entire instruction (n=2; 20%). One participant commented that it would have been useful to receive smaller sections of the instructions while using the QPL (n=1; 10%).

## Perceived Usefulness

Perceived usefulness was measured in terms of satisfaction with the content of the Web-based health information tools and intention to use the Web-based health information tools. Regarding satisfaction with the content of the Web-based health information tools, we identified 3 categories: (1) satisfaction with information modality, (2) information preferences, and (3) satisfaction with information comprehensibility.

### *Satisfaction With Information Modality*

Regarding the modality with which information was presented, the combination of text with a video was highly appreciated. Most participants commented that watching a video had added value after reading the text because it was difficult for them to process textual information only (n=11; 79%). Regarding illustrations, participants only found these useful when they clarified the text (n=6; 67%). One anatomical illustration on website 3 that used both Arabic and Latin numbers was difficult to understand. Illustrations that did not clarify the text, for example, a picture of a health care professional or a patient, received mixed comments. Some participants appreciated these illustrations (n=3; 13%), whereas other participants did not understand the reason why these illustrations were on the website and found these distracting (n=4; 17%).

### *Information Preferences*

When we asked participants what information they would search for after having received a cancer diagnosis and/or starting a cancer treatment, they indicated a need for the following information: (1) information about cancer type and/or treatment (n=14; 61%), (2) personally relevant information, for example, information on a specific treatment they would receive (n=10; 43%), and (3) contact information for hospitals and health care providers (n=6; 26%).

Website 3 offered testimonials of patients' experiences. Participants' opinions about these testimonials differed greatly: most (n=7; 78%) highly appreciated this information, whereas some had no need for this information at all (n=2; 22%).

One participant (17%) mentioned that information about alternative treatment options was missing on website 1. According to this participant, a health care provider should give a patient the choice to undergo a treatment or not, and (s)he preferred to retrieve not only information about the recommended treatment but also about the alternatives.

There were some comments on the amounts of questions and themes in QPLs 2 and 3. Some participants indicated that there were too many questions or themes in these QPLs, which demotivated them to use the tool (n=4; 24%). One participant was overwhelmed by the amount of questions:

> *when I see all these questions, I think that there are so many things I should worry about.*

### *Satisfaction With Comprehensibility*

Despite extensive instructions and example questions presented before using the tool, most participants mentioned that they had difficulties understanding the questions in the values clarification tool (n=21; 91%). The illustration to visualize, for example, a 2 of 100 chance that the tumor would come back, was not clear to the participants (n=6; 26%). Furthermore, participants had difficulties understanding the questions in which 2 paired scenarios were offered (n=16; 70%; see Materials—Decision Aid: Values Clarification Tool—for an explanation on the paired scenarios). The instructions were followed by example questions, which aimed to help the user understand the types of questions. However, the fact that the example questions were not cancer related was considered confusing by some participants (n=3; 13%). Two participants (9%) commented that the text was easy to understand as no foreign languages or medical jargon was used.

The answer categories for the questions in which participants had to answer whether they had a preference for one scenario over the other were considered too ambiguous, as participants were asked to give their preference and to state how strong their preference was in 1 question (n=12; 52%). One participant commented that

> *it would have been easier to just answer whether one has a preference for one scenario over the other or whether one has no preference at all.*

Another participant, while thinking aloud, said:

> *I will just answer that I have no preference, because I do not understand this question.*

Other participants were also observed to answer with the "no preference" option (n=8; 35%).

Participants were bothered by the number and apparent similarity of questions (n=9; 39%). Two participants (9%) commented that it would take them too long to finish the tool and that it took too long before they came to the relevant questions. This is because the tool started with questions about each of the treatment consequences first, followed by questions on the combined consequences of the treatment. Concerning the instructions, these were perceived as containing too much text, although it was not clear for participants what they could expect. Participants mentioned that they would rather see the question while reading the instructions (n=7; 30%).

### *Intention to Use*

Most participants mentioned that receiving information about their disease and treatment at home was very valuable, as it was very difficult for them to remember all the information presented during consultations (n=16; 70%). For example, 1 participant said that receiving information after the consultation is very useful as one can be too emotional to process information during the medical encounter. However, some participants mentioned that they would not use these types of websites as they expect to receive information from interpersonal communication with their health care providers and printed materials distributed by



XSL•FO
RenderX

their health care providers. Another participant commented that they did not want any information at the time of diagnosis. However, the participant continued,

> *the added value of a website with information is that one can select the information that one needs at the moment one wants to have the information.*

Participants had various needs regarding the amount of information. Some participants indicated that they were overwhelmed by large amounts of information (n=9; 39%), whereas other participants had a need for as detailed information as possible (n=4; 17%). One website offered the possibility of expanding the text for certain topics. This function was greatly appreciated by participants with both high- and low-information preferences (n=3; 33%).

Participants mentioned that the questions in the QPLs were useful for them and would help them to ask questions to their health care provider that they would not have thought of themselves (n=17; 74%). One participant, for example, mentioned that

> *you do not know what to expect before you have the consultation with your health care provider. It is very useful to see a list of possible topics that can be discussed.*

Although most participants thought the QPLs were useful when preparing for consultation, 2 (9%) had doubts about actual usage during the consultation as they thought that the health care provider did not have time to answer all the questions. Two (9%) other participants considered preparing for a consultation by thinking of possible questions to be useful but would not use a Web-based tool for this as they are used to doing this by pen and paper. Some other participants commented that they would not use a QPL or would prepare questions for a consultation in another way, as they expect to receive the information they need from the health care provider (n=3; 13%).

Participants mentioned that they had difficulties understanding the aim of the values clarification tool. When the researcher explained the aim of the tool, some participants did mention that such a tool would be useful for them as they could understand that the topics in the values clarification tool were important to think about (n=11; 48%). However, 1 participant mentioned that this goal could have been achieved by asking just 1 question: “what is important in your life?” Another participant commented that the goal of the tool would also have been achieved simply by asking “what is most important for you: recurrence of the tumor or the side effects of the treatment?” Most participants would not use the tool themselves as they did not understand the questions (n=16; 70%). Another reason for not using the tool was because some participants preferred to discuss the issues presented in the values clarification tool with their health care provider rather than using a tool (n=7; 30%).

## Discussion

### Principal Findings

The aim of this study was to provide insight into the user experience with existing Web-based health information tools among older cancer patients. We evaluated 7 different Web-based health information tools in terms of usability (ie, navigation strategy and navigation problems) and perceived usefulness (ie, content evaluations and intention to use).

Regarding usability, we identified how older cancer patients navigate through a website and which navigation problems they encounter. Older cancer patients had difficulties navigating through websites that had complex structures (eg, multiple navigation bars). Moreover, some navigation problems were attributable to the layouts of the websites. For example, some buttons were too small to click on for older patients suffering from physical decline. In addition, the age-related problem of visual decline played a role in navigation problems due to layout in that older patients had difficulties distinguishing colors that had low levels of contrast. Regarding the content that was presented on the websites, we found that older patients appreciated it when information was presented in different modalities (ie, text combined with illustrations or video). However, this combination was mostly appreciated if it was used to clarify the text and less for aesthetic reasons. Next, we found that older cancer patients and their partners varied greatly in terms of the amount of information they wanted to receive. Some patients wanted to receive as much information as possible, whereas other patients wanted to receive less information or no information at all. This finding is consistent with literature that found that older patients do not always want complete information on their disease [25]. All patients appreciated a website for which there was a possibility to expand information so that they could select the information they wanted to receive themselves.

The great effort it took for older adults to digest large amounts of information is probably also the reason why they preferred to only read what is applicable to their own situation, without having to filter it from among general information. This is in line with previous research that suggests that people find it increasingly difficult to concentrate on relevant information as they get older [41] and that older patients read large amounts of text when available [42].

Regarding the perceived usefulness of the Web-based health information tools, older adults overall indicated willingness to use both the health information websites and the preparatory tools. Reasons for not using the tools were that they would rather receive or discuss the information with a health care provider, that they preferred to receive offline information, or that they did not understand the content, which was the case for the values clarification tool. Similar results were found in usability testing of a comparable DA for chronic obstructive pulmonary disease patients [43].

### Strengths and Limitations

Although our participants were cancer patients or survivors and their partners, we asked participants to project themselves into



XSL•FO
RenderX

the hypothetical situation that they were just diagnosed with cancer or just about to receive a treatment. The use of such so-called "analog patients" is documented in meta-analysis as a valid method [44]. However, it may be more difficult even for cancer survivors to imagine the perceived usefulness of the system to a person newly diagnosed with cancer. To illustrate, not all information that was presented on the websites that we selected was applicable to the situations of the patients and their partners, which might have resulted in information that was not personally relevant. This possibly resulted in participants that were not as committed as the intended users of the websites and tools would be. Furthermore, although the usability problems of newly diagnosed patients might be similar to those of our analog patients, newly diagnosed patients may be more upset by usability problems that would make them unsure of whether the information they found applied to them or whether the decision they reached with the aid of the tool was the right one. This might affect the interpretation of the results and indicates that we must take even small usability problems very seriously. Moreover, prototypes of newly developed Web-based health information tools for older people should also be tested among recently diagnosed patients and partners.

We observed a difference between self-reported data and our observational data regarding the self-reported ease of finding information and the observed difficulty with actually finding the requested information, which points to the importance of using both self-reported data and observational data in user experience research. A possible explanation for this discrepancy is that participants may have given a socially desirable answer as the researcher was sitting next to the participant, although the researcher explained beforehand that the goal of the study was not to test the skills of the participant but the usability of the website.

### Comparison With Prior Work and Practical Implications

Previous guidelines focused on usability aspects of Web-based health information tools for older people, whereas this study also provides insights into perceived usefulness. Regarding usability, our study confirms some of the existing recommendations, refutes others, and suggests recommendations that are not mentioned in the existing guidelines. As Internet experience is increasing rapidly among older adults, some prior recommendations are no longer applicable to the current generation of older people. For example, our study showed that older website users can easily navigate through a pull-down menu—a nonstatic navigational element, whereas Pernice and Nielsen [9] more than a decade ago found that older Internet users had difficulties using these. The same authors [9] advise against scrolling down in a webpage. However, this study shows that most older users have no problems in doing so anymore.

The findings of this study confirm other existing recommendations. First, participants still had difficulties in reading small font sizes. It is important that websites designed for older users have adequate font sizes by default as participants were not able to find the button to increase font size. Second, similar to what we found in our study, Pernice and Nielsen [9] described that older users clicked on the back button in the browser to return to the homepage. Older users in our study were also not aware that clicking on the company logo would lead them back to the homepage. The recommendation to add a link called "Home" on all website pages except on the homepage and preferably in the horizontal navigation bar is therefore still applicable. Third, Pernice and Nielsen [9] recommended leaving space between links and to make the immediate area surrounding the link part of the link as older users have more difficulties with accurately clicking on small targets. This result is confirmed in our study, in which we found older users to be confused when they clicked on the wrong link or button or when nothing happened after misclicking the link.

Pernice and Nielsen [9] recommend presenting informational messages, including error messages, clearly and in a nonthreatening way. Although error messages were not common on the websites we tested in this study, we noticed that older users react in a confused or anxious manner when a pop-up unexpectedly shows up. Even when the pop-up is not an error message, participants interpreted it as such, which made them anxious. Another recommendation that was not found in the existing guidelines but that we would like to add is based on our finding that older users focus on the main text on a website instead of on navigational elements such as navigation bars. We therefore recommend presenting navigational elements in the center of the homepage, which will help older users immediately make a navigational choice without being distracted by possible irrelevant information. We also recommend avoiding large amounts of main text on the homepage and to display options on one page. For instance, if a clear overview with options is provided first, users can make a conscious choice regarding which information they want to read and click on the link or button with information that is relevant to them. To satisfy both users who prefer detailed information and users who want to read only key information, give text the ability to "pull out" for users who want to read more detailed information. This was highly appreciated by both groups on website 2. Make sure to use static menus and to not use more than 1 layer for pull-out menus to avoid users getting lost in the website.

Finally, this study builds on the existing guidelines in terms of providing insights regarding how to incorporate preparatory tools such as QPLs and values clarification tools. Based on the results of the think-aloud observations, we recommend providing clear instructions on how these tools can be used that are also available when using the tool. It is also recommended to limit the number of questions and themes in QPLs to a maximum of 20 predefined questions per QPL, to make 1 question visible at a time and to provide the possibility of adding additional personal questions. To be able to provide the user with a personal overview of all selected questions in order of priority, the option to add or to not add a selected question to this personal list (QPL) or answer (values clarification tool) should be provided, as well as the ability to prioritize the importance of questions, for example, by asking the user to indicate whether the topic or question is "not important (0)" to discuss, "rather important (1)," "important (2)," or "extremely important (3)." The ability to store this personal list, to print it out, or to email it should be incorporated into the tool. Textbox 2 gives a



XSL•FO
RenderX

summary of recommendations that can be derived from this study and previous studies.

One of the benefits of Web-based health information tools compared with traditional health information sources (eg, printed patient leaflets) is that information can be tailored to meet the individual preferences of individual patients [45]. Moreover, tailored information has been found to be processed more deeply, contain less redundant information, and is perceived more positively by users [46]. Although these results were found in a different context and in a younger sample, we expect that these benefits could also apply to older cancer patients as our results confirm that older cancer patients vary greatly in terms of their information preferences (ie, the amount and type of information they want to receive), literally evidenced by comments regarding the need to receive information that is personally relevant. Tailoring information according to patients' preferences would therefore make information more personally relevant, allow deeper information processing [47], and would contain less redundant information, which could be particularly beneficial to older patients considering age-related cognitive limitations [48].

Although the text of the values clarification tool was often perceived as too difficult to understand, participants thought that the goal of the values clarification tool (ie, a tool that would support them in thinking about which treatment consequences are most important for them) would be very useful. This is in line with a study [49] in which it was found that patients perceive such a values clarification tool as useful. A Web-based values clarification tool should therefore offer text or questions that are easy to understand and that prompt them to start thinking about their preferences. For example, a QPL consisting of 3 simple questions (ie, "what are my options?," "what are the possible benefits and harms of those options?," and "how likely are the benefits and harms of each option to occur?") has been designed [50]. The authors found that health care providers took patient preferences concerning treatment options into consideration after patients asked these 3 questions.

All Web-based health information tools were easier to use for older patients when they were provided with short instructions during use. Instructions that were given before Web-based health information tools were not remembered, if read at all. Short instructions should be provided while using the tool and should only apply to the specific function that is used at that time.



XSL•FO
RenderX

**Textbox 2.** Recommendations for the development of Web-based health information tools for older patients.

**General guidelines**

Recommendations based on current think-aloud observations:

- older people often use tablets: it is important that the site is suitable for a tablet;
- provide the ability to print information or to save as a PDF or send to email.

Recommendation based on both the current think-aloud observations and literature:

login must be simple.

Recommendation based on prior guidelines or literature:

ensure that the home page loads quickly.

**Access to information (structure and navigation)**

Recommendation based on current think-aloud observations:

avoid large amounts of information on a page. If possible, display options on 1 page, for example, first provide an overview with options, and then (after visitors choose what information they wish to read) the relevant information.

Recommendations based on both the current think-aloud observations and literature:

the design must focus on easy-to-use navigation tools:

- show a navigation bar on every page on the same place. (top)
- use a prominent homepage button on each page.

hyperlinks must be distinguishable from other text and easy to click on (eg, not too close to other text). Change the color if a link is clicked.

Recommendation based on prior guidelines or literature:

make sure that links go directly to the content.

**Information (text, illustrations, video, multimedia)**

Recommendations based on current think-aloud observations:

- allow text to "pull out" for users who want to read more detailed information (see eg,: http://www.avl.nl/behandelingen/chirurgie-bij-dikke-darmkanker/ under "What is going to happen");
- provide a clear explanation of illustrations: what exactly is there to see?

Recommendations based on both the current think-aloud observations and literature:

- combine strategies (audiovisual and text) so that older people have a choice. This is important because older people often want to control the pace at which they obtain information (which may be less possible with audiovisual information);
- a combination of personalization and audiovisual information enhances information memorization;
- adding images to a website improves satisfaction with the attractiveness of a website. Illustrations that explain the text are considered most useful. The images need to be carefully tested in the target group;
- large, readable font (so that a button to enlarge text is not required);
- text in a contrasting color background;
- write for users; do not use difficult language or technical terms;

Recommendations based on prior guidelines or literature:

- a website with a personal video (information provided by a patient) and text can increase satisfaction with the website;
- terms such as senior, older, or age-related terms can be used on the site. However, do not use stereotypes or patronizing text.

**Usability**

Recommendation based on both the current think-aloud observations and literature:

a search function is rarely used by the elderly. If this function is installed:

- make the search field very clear (eg, put the word "Search" clearly in front of open fields where users can search);
- the search engine should be easy to use and also work when punctuation is used;
- always repeat the search terms clearly above the search results;



XSL•FO
RenderX

- make sure the results are visible on the page without scrolling.
- buttons and other interactive objects must be easily clickable;
- error messages should be understandable and visible;
- when a pop-up is used, ensure that all information fits into the screen so that users do not have to scroll.

Recommendation based on prior guidelines or literature:

use static menus (no pull-down menus or other moving elements).

**If something needs to be filled in a form on the website:**

Recommendation based on prior guidelines or literature:

if you are asked for a date, use this format:

- select month by means of a drop-down list of months in chronological order;
- type date into an empty field;
- type year in 4-digit format in an empty field.
- if there are errors in a form, accept all the correct information and show users only the fields that need to be changed. Explain what the user should do to correct the error at the top of the form;
- ask users not to enter a salutation. Use a drop-down list if this information is required.

**Preparatory tools (QPLs or values clarification tools)**

Recommendations based on current think-aloud observations:

- provide clear instructions, which are also available when using the tool;
- limit the number of questions and themes in QPLs (up to 20 questions per topic);
- make 1 question at a time visible with the ability to add or not add the question to a personal list (QPL) or answer (values clarification tool);
- give an overview of all selected questions in order of priority and the option to add additional (personal) questions;
- provide the ability to store the list, print it out, or email it.

## Directions for Future Research

In addition to using existing guidelines for website development for older adults in general, our study shows the importance of taking the specific target group, in this case, older cancer patients, into consideration, as this group differs from a more general older target group. Future studies should investigate the user experience of other older patient groups as patients with other diseases might have different information needs or Web-based health information tools might have other functions such as medication reminders for patients with chronic diseases. Next, as the Netherlands is one of the countries with the highest Internet access among adults aged 65 years and older [51], future research is needed in countries where there are lower levels of Internet access.

Previous research concluded that QPLs can improve communication and psychological and cognitive outcomes in cancer patients (see [4] for a systematic review of the literature), and this was also found for older cancer patients [52]. This suggests that QPLs are useful tools to be developed and implemented for various diagnostic tests and treatments in cancer care. Although most participants considered a QPL to be a highly useful tool, this was not true for every older cancer patient. We identified certain reasons why older cancer patients would not intend to use a (Web-based) QPL, such as a preference for paper and pen and relying solely on interpersonal communication during consultation. The impression that the health care provider would not have time to answer the questions was also mentioned as a barrier, which is in line with previous research identifying barriers that patients have when discussing certain topics during consultations [53]. We therefore recommend that QPLs should not contain a large number of questions and should prioritize questions so that patients can ask their most important questions first, without increasing the consultation time (see Practical Implications). Future research should further investigate barriers for using Web-based health information tools such as QPLs.

The values clarification tool was also designed to be used to prepare patients for their consultation with their health care provider and to support the conversation about the weighing of benefits and harms of treatment. Participants indicated that the number of questions used in the values clarification tool was too extensive and that the importance of the outcomes could have been assessed by asking them in one direct question. However, the purpose of the adaptive conjoint analysis is that the relative values are assessed, that is, the importance of an outcome in relation to the other outcomes. Participant comments indicated that they would rather discuss the advantages and disadvantages of a treatment with their health care provider instead of using the tool, which might differ when participants actually used the tool in combination with interpersonal communications with their health care provider. As the values



XSL•FO
RenderX

clarification tool is designed and had been used in the context of a clinical study only in which the tool was combined with consultations with health care providers, the comments of the participants in this think-aloud study were that they would rather discuss these benefits and harms of treatment with their health care provider are therefore not unexpected. A recent literature review on the effectiveness of decision aids for older adults indicated that patient outcomes seemed to be better when participants received the decision aid from their clinician during the consultation than when it was delivered by a researcher before the consultation [54,55]. This suggests that decision aids might be particularly useful for older adults when successfully integrated with interpersonal communication during the consultation. However, only 2 studies in which the decision aid was delivered during the consultation were included in the review [56]. The same might hold true for the QPLs. Future studies should therefore examine the added value of these tools when offered by the health care provider during the consultation.

## Conclusions

This study shows how older cancer patients use and evaluate Web-based health information tools. Older cancer patients are fully able to use Web-based health information tools and perceive these tools as highly useful in their search for health information and to prepare for interpersonal communication with their health care providers. However, older patients experienced navigational problems that can hinder optimal user experience with these tools. This study unmasked these navigation problems along with specific user preferences. We used our results to propose improvements for the design of Web-based health information tools for optimal user experience among older patients.

## Acknowledgments

The Dutch Cancer Society (KWF; grant numbers UVA 2010-4740 and UVA 2013-6460) and the Amsterdam School of Communication Research/ASCoR, University of Amsterdam funded this study. KWF and ASCoR management were not involved in the study design, data collection, data analysis, report writing, and decision to submit the manuscript for publication. The authors would like to thank Nadine Bol for her assistance in collecting the data. Next, the authors would like to thank Marc Jacobs and Marleen Kunneman for making the QPL “Mijn Vragenhulp” and the values clarification tool, respectively, available for this study. The authors would like to thank Felicitas van Alphen for her work as a second coder.

## Conflicts of Interest

None declared

## Multimedia Appendix 1

Screenshot of QPL 1.

[PDF File (Adobe PDF File), 108KB - jmir_v18i7e208_app1.pdf ]

## Multimedia Appendix 2

Screenshots of values clarification tool.

[PNG File, 101KB - jmir_v18i7e208_app2.png ]

## References

1. 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]
2. Van de Belt TH, Engelen Lucien LJ, Berben SA, Teerenstra S, Samsom M, Schoonhoven L. Internet and social media for health-related information and communication in health care: preferences of the Dutch general population. J Med Internet Res 2013;15(10):e220 [FREE Full text] [doi: 10.2196/jmir.2607] [Medline: 24088272]
3. Murray E, Burns J, See TS, Lai R, Nazareth I. Interactive Health Communication Applications for people with chronic disease. Cochrane Database Syst Rev 2005(4):CD004274. [doi: 10.1002/14651858.CD004274.pub4] [Medline: 16235356]
4. Brandes K, Linn AJ, Butow PN, van Weert JCM. The characteristics and effectiveness of Question Prompt List interventions in oncology: a systematic review of the literature. Psychooncology 2015 Mar;24(3):245-252. [doi: 10.1002/pon.3637] [Medline: 25082386]
5. Pieterse AH, Berkers F, Baas-Thijssen MC, Marijnen CA, Stiggelbout AM. Adaptive Conjoint Analysis as individual preference assessment tool: feasibility through the internet and reliability of preferences. Patient Educ Couns 2010 Feb;78(2):224-233. [doi: 10.1016/j.pec.2009.05.020] [Medline: 19581069]
6. Barnett K, Mercer SW, Norbury M, Watt G, Wyke S, Guthrie B. Epidemiology of multimorbidity and implications for health care, research, and medical education: a cross-sectional study. Lancet 2012 Jul 7;380(9836):37-43 [FREE Full text] [doi: 10.1016/S0140-6736(12)60240-2] [Medline: 22579043]



XSL•FO
RenderX

7. Bolle S, van Weert JC, Daams JG, Loos EF, de Haes JC, Smets EM. Online Health Information Tool Effectiveness for Older Patients: A Systematic Review of the Literature. J Health Commun 2015;20(9):1067-1083. [doi: 10.1080/10810730.2015.1018637] [Medline: 26165846]
8. Bolle S, van Weert JC, Smets EM, Loos EF. Lack of Development and Usability Descriptions in Evaluation Reports on Online Health Information Tools for Older Patients. In: Zhou J, Salvendy G, editors. Human aspects of IT for the aged population: design for everyday life; first international conference, ITAP 2015, held as part of HCI International 2015, Los Angeles, CA, USA, August 2 - 7, 2015; proceedings, part II (Lecture Notes in Computer Science, 9194). Cham, Switzerland: Springer International Publishing; 2015:27-37.
9. Pernice K, Nielsen J. Web usability for senior citizens. Design guidelines based on usability studies with people age 65 and older. Fremont, CA: Nielsen Norman Group; 2002.
10. Loos EF, Romano Bergstrom J. Older adults. In: Romano Bergstrom J, Schall AJ, editors. Eye Tracking in User Experience Design. Waltham, MA: Elsevier; 2014:313-329.
11. Becker SA. A study of web usability for older adults seeking online health resources. ACM Trans Comput-Hum Interact 2004 Dec 01;11(4):387-406. [doi: 10.1145/1035575.1035578]
12. Nayak L, Priest L, Stuart-Hamilton I, White A. Website design attributes for retrieving health information by older adults: an application of architectural criteria. Univ Access Inf Soc 2006 Apr 12;5(2):170-179. [doi: 10.1007/s10209-006-0029-9]
13. Czaja SJ, Charness N, Fisk AD, Hertzog C, Nair SN, Rogers WA, et al. Factors predicting the use of technology: findings from the Center for Research and Education on Aging and Technology Enhancement (CREATE). Psychol Aging 2006 Jun;21(2):333-352 [FREE Full text] [doi: 10.1037/0882-7974.21.2.333] [Medline: 16768579]
14. Crutzen R. From eHealth technologies to interventions. J Med Internet Res 2012;14(3):e93 [FREE Full text] [doi: 10.2196/jmir.2050] [Medline: 22738742]
15. van Gemert-Pijnen JE, Nijland N, van LM, 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]
16. Burns P, Jones SC, Iverson D, Caputi P. Usability testing of AsthmaWise with older adults. Comput Inform Nurs 2013 May;31(5):219-26; quiz 227. [doi: 10.1097/NXN.0b013e31828a0dc2] [Medline: 23542306]
17. Hinchliffe A, Mummery WK. Applying usability testing techniques to improve a health promotion website. Health Promot J Austr 2008 Apr;19(1):29-35. [Medline: 18481929]
18. Yen P, Bakken S. Review of health information technology usability study methodologies. J Am Med Inform Assoc 2012 Jun;19(3):413-422 [FREE Full text] [doi: 10.1136/amiajnl-2010-000020] [Medline: 21828224]
19. Law E, Roto V, Hassenzahl M, Vermeeren A, Kort J. Understanding, scoping and defining user experience: a survey approach. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. 2009 Presented at: SIGCHI Conference on Human Factors in Computing Systems; 4-9 April, 2009; Boston. [doi: 10.1145/1518701.1518813]
20. Davis JF. A technology acceptance model for empirically testing new end-user information systems: Theory and results [dissertation]. Cambridge, MA: Sloan School of Management, Massachusetts Institute of Technology; 1986.
21. Lee Y, Kozar KA, Larsen KR. The technology acceptance model: Past, present, and future. Communications of the Association for information systems 2003;12(1):50.
22. Davis FD. Perceived Usefulness, Perceived Ease of Use, and User Acceptance of Information Technology. MIS Quarterly 1989 Sep;13(3):319-340. [doi: 10.2307/249008]
23. van Waes L. Thinking aloud as a method for testing the usability of Websites: the influence of task variation on the evaluation of hypertext. IEEE Trans Profess Commun 2000;43(3):279-291. [doi: 10.1109/47.867944]
24. Taha J, Sharit J, Czaja S. Use of and satisfaction with sources of health information among older Internet users and nonusers. Gerontologist 2009 Oct;49(5):663-673. [doi: 10.1093/geront/gnp058] [Medline: 19741112]
25. Giacalone A, Blandino M, Talamini R, Bortolus R, Spazzapan S, Valentini M, et al. What elderly cancer patients want to know? Differences among elderly and young patients. Psychooncology 2007 Apr;16(4):365-370. [doi: 10.1002/pon.1065] [Medline: 16921547]
26. Jaspers MW, Steen T, van den Bos C, Geenen M. The think aloud method: a guide to user interface design. Int J Med Inform 2004 Nov;73(11-12):781-795. [doi: 10.1016/j.ijmedinf.2004.08.003] [Medline: 15491929]
27. Bunz UK. Usability and Gratifications -- Towards a Website Analysis Model. In: International Journal of Human-Computer Studies. 2001 Presented at: Annual Meeting of the National Communication Association; November 1-4, 2001; Atlanta, GA.
28. Nielsen J. Estimating the number of subjects needed for a thinking aloud test. International Journal of Human-Computer Studies 1994 Sep;41(3):385-397. [doi: 10.1006/ijhc.1994.1065]
29. Virzi R. Refining the test phase of usability evaluation: How many subjects is enough? Human Factors: The Journal of the Human Factors and Ergonomics Society 1992;34(4):457-468. [doi: 10.1177/001872089203400407]
30. Jaspers MW. A comparison of usability methods for testing interactive health technologies: methodological aspects and empirical evidence. Int J Med Inform 2009 May;78(5):340-353. [doi: 10.1016/j.ijmedinf.2008.10.002] [Medline: 19046928]




XSL•FO
RenderX

31. Aitken LM, Marshall A, Elliott R, McKinley S. Comparison of 'think aloud' and observation as data collection methods in the study of decision making regarding sedation in intensive care patients. Int J Nurs Stud 2011 Mar;48(3):318-325. [doi: 10.1016/j.ijnurstu.2010.07.014] [Medline: 21056417]
32. Nikolaus S, Bode C, Taal E, Vonkeman HE, Glas CA, van de Laar MA. Acceptance of New Technology: A Usability Test of a Computerized Adaptive Test for Fatigue in Rheumatoid Arthritis. JMIR Hum Factors 2014;1(1):e4 [FREE Full text] [doi: 10.2196/humanfactors.3424] [Medline: 27025404]
33. Morahan-Martin JM. How internet users find, evaluate, and use online health information: a cross-cultural review. Cyberpsychol Behav 2004 Oct;7(5):497-510. [Medline: 15667044]
34. Bolle S, Muusses LD, Smets EM, Loos EF, van Weert JC. Een onderzoek naar de publieke kennis, perceptie van bijwerkingen, informatiebehoeften en het gebruik van informatiebronnen met betrekking tot chemotherapie. Amsterdam: University of Amsterdam; 2012.
35. Chemo & nu. Wegwijs in chemotherapie URL: http://www.chemotherapie.nl/ [accessed 2016-02-08] [WebCite Cache ID 6f89mrTo8]
36. Antoni van Leeuwenhoek. Nederlands Kanker Instituut URL: http://www.avl.nl/ [accessed 2016-02-08] [WebCite Cache ID 6f89Re0h7]
37. AMC. GIOCA: expertisecentrum voor gastro-intestinale kanker URL: https://www.amc.nl/web/Zorg/Expertisecentrum-Gioca/Gioca/Expertisecentrum.htm [accessed 2016-02-08] [WebCite Cache ID 6f89getQR]
38. Chemo & nu. Een goed gesprek URL: http://www.chemotherapie.nl/media/rtf/chemo_eengoedgesprek.pdf [accessed 2016-06-11] [WebCite Cache ID 6iBRvENNw]
39. B-bewust. Beter voorbereid URL: http://www.b-bewust.nl/ [accessed 2016-02-08] [WebCite Cache ID 6f89qVTUY]
40. Pieterse AH, de Vries M, Kunneman M, Stiggelbout AM, Feldman-Stewart D. Theory-informed design of values clarification methods: a cognitive psychological perspective on patient health-related decision making. Soc Sci Med 2013 Jan;77:156-163. [doi: 10.1016/j.socscimed.2012.11.020] [Medline: 23219164]
41. Kray J, Lindenberger U. Adult age differences in task switching. Psychol Aging 2000 Mar;15(1):126-147. [Medline: 10755295]
42. Tullis TS. Older adults and the web: lessons learned from eye-tracking. In: Universal Acess in Human Computer Interaction. Coping with Diversity. Cham, Switzerland: Springer; 2007:1030-1039.
43. Uhler LM, Pérez Figueroa RE, Dickson M, McCullagh L, Kushniruk A, Monkman H, et al. InformedTogether: Usability Evaluation of a Web-Based Decision Aid to Facilitate Shared Advance Care Planning for Severe Chronic Obstructive Pulmonary Disease. JMIR Human Factors 2015 Feb 25;2(1):e2. [doi: 10.2196/humanfactors.3842]
44. van Vliet LM, van der Wall E, Albada A, Spreeuwenberg PM, Verheul W, Bensing JM. The Validity of Using Analogue Patients in Practitioner–Patient Communication Research: Systematic Review and Meta-Analysis. J Gen Intern Med 2012 Jun 15;27(11):1528-1543. [doi: 10.1007/s11606-012-2111-8]
45. Cline RJ, Haynes KM. Consumer health information seeking on the Internet: the state of the art. Health Educ Res 2001 Dec;16(6):671-692 [FREE Full text] [Medline: 11780707]
46. Brug J, Oenema A, Campbell M. Past, present, and future of computer-tailored nutrition education. Am J Clin Nutr 2003 Apr;77(4 Suppl):1028S-1034S [FREE Full text] [Medline: 12663313]
47. Hawkins RP, Kreuter M, Resnicow K, Fishbein M, Dijkstra A. Understanding tailoring in communicating about health. Health Educ Res 2008 Jun;23(3):454-466 [FREE Full text] [doi: 10.1093/her/cyn004] [Medline: 18349033]
48. Brown SC, Park DC. Theoretical models of cognitive aging and implications for translational research in medicine. Gerontologist 2003 Mar;43 Spec No 1:57-67. [Medline: 12637690]
49. Pieterse AH, Stiggelbout AM, Baas-Thijssen MCM, van de Velde CJH, Marijnen CAM. Benefit from preoperative radiotherapy in rectal cancer treatment: disease-free patients' and oncologists' preferences. Br J Cancer 2007 Sep 17;97(6):717-724 [FREE Full text] [doi: 10.1038/sj.bjc.6603954] [Medline: 17848910]
50. Shepherd HL, Barratt A, Trevena LJ, McGeechan K, Carey K, Epstein RM, et al. Three questions that patients can ask to improve the quality of information physicians give about treatment options: a cross-over trial. Patient Educ Couns 2011 Sep;84(3):379-385. [doi: 10.1016/j.pec.2011.07.022] [Medline: 21831558]
51. Statistics N. Statline CBS. 2014. ICT gebruik van personen naar persoonskenmerken [ICT use by persons subject to personal characteristics] URL: http://statline.cbs.nl/Statweb/publication/?DM=SLNL&PA=71098ned&D1=33,55-59,69-72,93-95,97-100,102-104,106-109,112-113,119,124-133&D2=0,4-6&D3=0,l&HDR=G2,G1&STB=T&VW=T [WebCite Cache ID 6f8D3903k]
52. van Weert JC, Jansen J, Spreeuwenberg PM, van Dulmen S, Bensing JM. Effects of communication skills training and a Question Prompt Sheet to improve communication with older cancer patients: A randomized controlled trial. Critical Reviews in Oncology/Hematology 2011 Oct;80(1):145-159. [doi: 10.1016/j.critrevonc.2010.10.010]
53. Brandes K, Linn AJ, Smit EG, van Weert JCM. Patients' reports of barriers to expressing concerns during cancer consultations. Patient Educ Couns 2015 Mar;98(3):317-322. [doi: 10.1016/j.pec.2014.11.021] [Medline: 25499005]
54. Jones LA, Weymiller AJ, Shah N, Bryant SC, Christianson TJ, Guyatt GH, et al. Should clinicians deliver decision aids? Further exploration of the statin choice randomized trial results. Med Decis Making 2009;29(4):468-474. [doi: 10.1177/0272989X09333120] [Medline: 19605885]




XSL•FO
RenderX

55. Weymiller AJ, Montori VM, Jones LA, Gafni A, Guyatt GH, Bryant SC, et al. Helping patients with type 2 diabetes mellitus make treatment decisions: statin choice randomized trial. Arch Intern Med 2007 May 28;167(10):1076-1082. [doi: 10.1001/archinte.167.10.1076] [Medline: 17533211]
56. van Weert JCM, van Munster BC, Sanders R, Spijker R, Hooft L, Jansen J. Decision aids to help older people make health decisions: a systematic review and meta-analysis. BMC Med Inform Decis Mak 2016;16:45 [FREE Full text] [doi: 10.1186/s12911-016-0281-8] [Medline: 27098100]

## Abbreviations

**QPL:** question prompt list

*Edited by G Eysenbach; submitted 07.02.16; peer-reviewed by S Medlock, P Burns; comments to author 19.04.16; revised version received 13.06.16; accepted 28.06.16; published 25.07.16.*

*Please cite as:*
*Bolle S, Romijn G, Smets EMA, Loos EF, Kunneman M, van Weert JCM*
*Older Cancer Patients' User Experiences With Web-Based Health Information Tools: A Think-Aloud Study*
*J Med Internet Res 2016;18(7):e208*
*URL: http://www.jmir.org/2016/7/e208/*
*doi:10.2196/jmir.5618*
*PMID:27457709*

©Sifra Bolle, Geke Romijn, Ellen M A Smets, Eugene F Loos, Marleen Kunneman, Julia C M van Weert. Originally published in the Journal of Medical Internet Research (http://www.jmir.org), 25.07.2016. This is an open-access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/2.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on http://www.jmir.org/, as well as this copyright and license information must be included.



XSL•FO
RenderX

Original Paper

# Social Annotation Valence: The Impact on Online Informed Consent Beliefs and Behavior

Martina Balestra[1]; Orit Shaer[2], PhD; Johanna Okerlund[2]; Lauren Westendorf[2]; Madeleine Ball[3], PhD; Oded Nov[1], PhD

[1]Department of Technology Management & Innovation, Tandon School of Engineering, New York University, Brooklyn, NY, United States

[2]HCI Lab, Wellesley College, Wellesley, MA, United States

[3]personalgenomes.org, New York, NY, United States

**Corresponding Author:**
Martina Balestra
Department of Technology Management & Innovation
Tandon School of Engineering
New York University
5 MetroTech Center
Brooklyn, NY, 11201
United States
Phone: 1 415 260 9758
Fax: 1 646 997 3136
Email: mb5758@nyu.edu

## Abstract

**Background:** Social media, mobile and wearable technology, and connected devices have significantly expanded the opportunities for conducting biomedical research online. Electronic consent to collecting such data, however, poses new challenges when contrasted to traditional consent processes. It reduces the participant-researcher dialogue but provides an opportunity for the consent deliberation process to move from solitary to social settings. In this research, we propose that social annotations, embedded in the consent form, can help prospective participants deliberate on the research and the organization behind it in ways that traditional consent forms cannot. Furthermore, we examine the role of the comments' valence on prospective participants' beliefs and behavior.

**Objective:** This study focuses specifically on the influence of annotations' valence on participants' perceptions and behaviors surrounding online consent for biomedical research. We hope to shed light on how social annotation can be incorporated into digitally mediated consent forms responsibly and effectively.

**Methods:** In this controlled between-subjects experiment, participants were presented with an online consent form for a personal genomics study that contained social annotations embedded in its margins. Individuals were randomly assigned to view the consent form with positive-, negative-, or mixed-valence comments beside the text of the consent form. We compared participants' perceptions of being informed and having understood the material, their trust in the organization seeking the consent, and their actual consent across conditions.

**Results:** We find that comment valence has a marginally significant main effect on participants' perception of being informed ($F_2$=2.40, $P$=.07); specifically, participants in the positive condition (mean 4.17, SD 0.94) felt less informed than those in the mixed condition (mean 4.50, SD 0.69, $P$=.09). Comment valence also had a marginal main effect on the extent to which participants reported trusting the organization ($F_2$=2.566, $P$=.08). Participants in the negative condition (mean 3.59, SD 1.14) were marginally less trusting than participants exposed to the positive condition (mean 4.02, SD 0.90, $P$=.06). Finally, we found that consent rate did not differ across comment valence conditions; however, participants who spent less time studying the consent form were more likely to consent when they were exposed to positive-valence comments.

**Conclusions:** This work explores the effects of adding a computer-mediated social dimension, which inherently contains human emotions and opinions, to the consent deliberation process. We proposed that augmenting the consent deliberation process to incorporate multiple voices can enable individuals to capitalize on the knowledge of others, which brings to light questions, problems, and concerns they may not have considered on their own. We found that consent forms containing positive valence annotations are likely to lead participants to feel less informed and simultaneously more trusting of the organization seeking



XSL•FO
RenderX

consent. In certain cases where participants spent little time considering the content of the consent form, participants exposed to positive valence annotations were even more likely to consent to the study. We suggest that these findings represent important considerations for the design of future electronic informed consent mechanisms.

***(J Med Internet Res 2016;18(7):e197)*** doi:10.2196/jmir.5662

**KEYWORDS**

consent forms; decision support systems; social tagging systems; informed consent; ethics

## Introduction

Social media, mobile and wearable technology, and connected devices have significantly expanded the opportunities for conducting research online. Already recognized as a rich resource for psychological and social research [1], biomedical research is taking increasing interest in these digital methods. Apple's launch of ResearchKit in April 2015 provides an example of a tool created specifically to facilitate biomedical research through online processes and interactions [2]. The reduced barrier to entry for participation in online biomedical research and the sensitivity of the resultant data highlight the importance of informed online consent processes and require us to reevaluate their effectiveness and potential to enhance the consent deliberation process in this new context.

Electronic consent poses new challenges when contrasted to traditional consent processes. Whereas individuals were formerly able to engage with a professional in additional face-to-face dialogue, potential online research participants have fewer opportunities to ask questions and express their concerns in real time. Furthermore, the use of certain presentation techniques and design interventions may influence an individual's decision to participate [3,4], raising concerns regarding voluntariness. In response to these and other concerns, federal agencies are drafting guidelines for electronic consent [5].

While electronic consent can reduce the participant-researcher dialogue, the online environment allows the consent deliberation process to move from solitary to social settings. A computer-supported social environment could enable individuals deliberating on their consent decision to connect with each other, share information, formulate and evaluate different perspectives, and ultimately understand the risks and benefits of the research beyond the scope of one-on-one dialogue with a research staff member.

In a previous study [6], we hypothesized that incorporating user-generated social annotations into online consent forms with complex content would allow individuals to benefit and learn from others' perspectives, knowledge, and ideas by encouraging discussion and helping to focus attention on the issues that users find important. We designed such a tool and evaluated it compared to a control condition of an online consent form with no social annotation. Specifically, we compared participants' perceptions of the extent to which they felt informed when they made their consent decisions, the extent to which they felt that they understood the content of the consent form, and the extent to which they trusted the organization seeking consent with the perceptions and beliefs of participants in the control condition. While the social annotation intervention did not influence the consent rate, we found that individuals exposed to social annotations in consent forms felt more informed compared to those exposed to traditional online consent forms, and furthermore, that the effect of exposure to social annotation was stronger among users who were less concerned about privacy. Interestingly, we also found that participants felt that they understood the consent form and trusted the organization more in the control condition than when exposed to social annotation. Taken together, the results indicated that social annotations can serve to highlight individuals' own limitations in comprehension and engage participants around the negative aspects of the consent form rather than the positive aspects, leading to lower levels of trust and perceived comprehension.

Following our first study, a number of questions remained concerning the extent to which annotations containing bias or emotional valence may influence users' deliberative processes and consent decisions, and the necessity of "policing" such information contributed by anonymous users in a high-risk context. User-generated content contains human emotion and bias by its very nature and can influence others: "…affect appears to influence what we notice, what we learn, what we remember, and ultimately the kinds of judgments and decisions we make" (p. 273) [7]. This study builds on and extends our previous research to understand the influence of annotations' valence on perceptions and behaviors surrounding online consent. In doing so, we hope to shed light on how social annotation can be incorporated into digitally mediated consent forms responsibly and effectively.

### Application Domain: Personal Genomics

Traditionally, medical genetic testing targeted individual loci and was performed for specific medical contexts (eg, when investigating a suspected genetic condition). A medical expert mediated the consent process for testing and returning results. A precipitous decline in the costs of genome-scale testing, however, has led to widespread access of personal genomic data. Several companies currently offer genome-scale testing services directly to consumers. Direct-to-consumer genetic testing (DTCGT) is a relatively new and developing online service that enables individuals to acquire genetic information without the mandatory involvement of a health care provider by sending a saliva sample to a DTCGT company at the cost of a few hundred dollars. DTCGT users are often asked to share their genetic and family history information with biomedical researchers who partner with the DTCGT provider. Genetic results, including traits, ancestry, and in some cases, health information, are reported using interactive online apps [8,9]. With DTCGT, computer-mediated consent and the presentation of results have become core aspects of giving individuals access



XSL•FO
RenderX

to their genome-scale test results. At the same time, these aspects raise concerns that policy makers as well as researchers attempt to address [5]. Many of the risks associated with digitally mediated genetic testing are related to data privacy: the technical limitations of keeping genomic data safe and secure, the possibilities for unintended public disclosure and identifiability if those records become public, the potential for genetic discrimination by the law, employers, or insurance agencies, and the handling and potential for misuse by research personnel. The consent form is responsible for communicating the gravity and significance of these risks and others to participants with varying degrees of knowledge of genomics and data privacy, as well as varying degrees of concern with privacy-related issues.

## Informed Consent

The decision to consent to participate in biomedical research is generally mediated by two main factors: participants' comprehension of the details of the study and their trust in the research organization [10]. Informed consent consists of four core tenets (ie, disclosure, comprehension, voluntariness, competence) and describes the process of educating individuals on a procedure so that they are able to make a well-reasoned decision about their voluntary agreement to participate [3,11]. The moral obligation of consent seekers is widely recognized as providing "those facts that all rational persons would want to know, namely, the various goods and evils that result from alternative modes of treatment, including severity and probability" [12]. Ubel and Lowenstein [13] suggest that this approach falls short of helping individuals make decisions that fit with their own values. They propose to find a way to combine medical facts with attributes and considerations that are relevant to participants with suspicions, hopes, fears, and anxieties. With this study, we assert that adding a computer-supported social aspect to the consent deliberation process means bringing in other perspectives on what "information" is valuable for informed consent.

## Consent Forms

Prior research on the design of consent forms has not yielded consistent results. Early studies on the design of consent forms focused on text readability [14,15]. Following the realization that readability does not necessarily relate to comprehension [16], research shifted to explore different ways to communicate the content of consent forms and other legal documents. Recent studies on consent form design focused predominantly on the impact of content structure, graphical enhancements, and multimedia on comprehension. Dresden and Levitt [17] demonstrated greater comprehension when a consent form was shortened to contain only details that the researchers believed most relevant to a potential participant. In a test comparing comprehension of a traditional consent form and a graphically enhanced form, however, Stiles et al [18] found no significant difference in the rate of comprehension. Murphy et al [19] showed a significant increase in consent form comprehension scores with a combination of restructured text, simplified vocabulary and sentence structure, and the use of illustrations to communicate key concepts. Dunn et al [20] found that the participants assigned to read a consent form formatted as a structured, computerized slideshow scored higher in comprehension tests than participants assigned to a traditional consent form condition. Other studies, however, show that replacing a traditional consent form with an interactive computer-based presentation does not result in consistent improvements in comprehension [21,22]. Multimedia interventions have used video to replace or complement textual consent forms, though comprehension tests have widely demonstrated that video has little effect on consent form comprehension [22,23].

## Social Annotation

Social annotations consist of three elements: the resource (ie, the text in question), the users, and the metadata created by the users. In a paper on the collective dynamics of social annotation, Catutto et al [24] define social annotation as "freely established associations between Web resources and metadata [keywords and descriptive labels, categories, ratings, comments and notes] performed by a community of Web users with little or no central coordination" (p. 10511) that captures the relevant collective knowledge of all users. Gao [25] asserts that access to this type of social annotation allows users to discuss content collaboratively and asynchronously, and presents evidence that there is more discussion that is more thoughtful, focused, and related to the text when users had access to social annotations. Further, Nelson et al [26] demonstrated substantial learning effects among participants in exploratory learning tasks who had access to social annotations during a controlled laboratory experiment. Within the context of consent forms, incorporating social information may allow individuals to benefit and learn from others' novel perspectives, knowledge, and ideas by encouraging discussion and helping focus attention on the issues they find important.

Cross and Sproull [27] argue that the value of social information is fundamental and not limited to the online environment. In a qualitative study of information relationships, the authors found that individuals tend to seek out relationships that support problem reformulation (in which others help to define or redefine dimensions of a problem not previously considered). In the context of the social consent form, Cross and Sproull's [27] findings show that individuals would perceive the information relationships embodied in social annotations as valuable resources for vetting the risks and benefits of participation.

Access to socially constructed information can impact the decisions an individual makes in areas ranging from consumer products [28] to travel [29] and security feature adoption [30]. Das et al [30] found that information exchanges on the topic of security tend to begin with an individual's desire to warn others of immediate or novel threats, or to acquire information useful for understanding a particular system or solving a problem. This suggests to us that participants would be motivated to use social annotations in the context of consent for biomedical research and that the decisions they make about consenting could be influenced in turn by the knowledge and experiences of others.

When user-contributed information is generated and added voluntarily to digitally mediated documents, they are not usually policed by a centralized authority [23] and therefore annotations



XSL•FO
RenderX

may contain inaccurate information or perceptions. Though Bernstein et al [31] used the social features of *Collabio* to show that the tags produced by users had a high degree of accuracy, they attributed this accuracy to social motivators that prevented serious misuse or off-topic tags. These social motivators may not necessarily exist in a context like medical research where anonymity is not only valued, but also legally mandated. Further, in the absence of personal identifiers, potential participants may perceive certain others as "experts," who are more valuable and more persuasive than others, where they might not necessarily be [32,33].

Any potential for false information can have significant impacts on prospective participants. An individual's ability to respond appropriately to a situation requires the ability to correctly interpret and react to incoming information, particularly in compliance-gaining settings [34]. The individual relying on socially constructed information may therefore be making decisions based on erroneous information or misplaced beliefs, which can impact not only the participant, but in cases like genomic research, also participants' ancestors and offspring.

### Message Valence and Social Annotation

Social annotations communicate both information and emotion: as a form of human communication they inherently carry information about the contributor's emotional state or judgment about the content [35]. One outcome of this is the development of an emotional connection with content that would otherwise be static or impersonal [36]. The utility of social annotations in the process and experience of deliberation, however, is not well understood.

Prior research on the influence of user-generated comment valence has largely been done in the context of consumer reviews. Chen and Xie [37] argue that consumer reviews generated by users based on their individual experiences can help subsequent customers find products matching their needs. They also assert that the information provided by the institution and user-generated content act as substitutes for each other, rather than complements, when the cost of the product is high and reviews are generated by novice reviewers.

Studies on text with affective dimensions suggest that positive and negative sentiment could lead to greater cognitive involvement in terms of attention as well as better memory of the text [38,39]. Smith and Petty [40] showed that message framing impacts the extent to which an individual processes the message. Specifically, they found that messages whose framing was unexpected led to more extensive message processing. The authors drew on Kahenman and Tversky's Prospect Theory [41] to define positive framing as the characterization of uncertain alternatives in terms of potential gains, and negative framing in terms of potential losses. The individual is engaged because the message is more salient. Applying prospect theory to persuasion, they also noted that negatively framed messages should be more persuasive than positively framed messages.

Messages evoking or communicating particular sentiments result in different forms of engagement with the message. Berger [42] found that content evoking particular sentiments can ultimately lead to higher levels of arousal resulting in higher rates of sharing. In a study on the relationship between blog sentiment and the volume of feedback, Dang-Xuan and Stieglitz [43] found that blog posts with negative and positive valence elicited significantly more comments compared with neutral or mixed valence blog posts. Affect, as an impetus for reaction, seems to exist in other contexts as well: online leadership in discussion forums appears to be positively correlated with the use of emotional valence in messages [44]. Furthermore, negative and positive valence messages do not necessarily produce the same outcomes: messages with positive valence tend to evoke a sense of community that encourages participation, whereas negative valence comments can result in more hostile and heated exchanges [45].

### Trust and Social Annotation

Beyond the effective and appropriate communication of information, previous research shows that trust plays a crucial role in the decision to disclose sensitive information online [46]. Similarly, trusting the physician or research organization plays a fundamental role in the decision to participate in medical research [47]. We view trust in the medical context as "the expectation that institutions and professionals will act in one's interests" (pg. 661); this view follows from [48]. In this context, trust consists of five dimensions: expectations about the research organization's competence, the extent to which the organization is concerned with their patient's welfare, the organization's control over decision making, the organization's management of confidential information, and the organization's openness in providing and receiving information [48]. In traditional consent-seeking procedures, the individual independently examines the information provided by the authors of a consent form along these five dimensions before making a decision about consent. By implementing social annotations, we enable prospective participants to capitalize on the experiences of others to discern trustworthiness and therefore add a social perspective to the user's development of trust in the organization seeking consent.

### Hypotheses

Drawing from the literature above, our research model is depicted in Figure 1. Independent variables include comment valence (listed on the left); dependent variables include consent, the extent to which users felt that their decision was informed, the extent to which users felt that they understood the material, and the extent to which they trust the organization offering the study (listed on the right). A measure of participants' concern for privacy-related issues in the digital environment served as an interaction term (listed on the bottom of the diagram). We also measured interactivity with the annotations to give us a more general idea of how participants used the annotations. The arrows denote the hypotheses addressed in this study.

Deliberating whether to participate in medical research can be a complex process, though individuals' decision-making abilities are limited [13]. Prior research has shown that in such scenarios, individuals tend to simplify these deliberations by ignoring large amounts of information while focusing on a subset of information relevant to their value system [18,41]. In the context of consent, we suggest that social annotations serve to connect individuals' value systems to the content of the consent form



XSL•FO
RenderX

in an explicit manner. Prospective participants are able to observe, identify with, and learn from the issues, questions, concerns, and emotions communicated by previous participants on topics relevant to their values, thus focusing their deliberations on these issues and improving the efficiency and effectiveness of their deliberative process. Nelson et al [26] showed that social annotations can be useful in helping individuals learn unfamiliar topics. Social annotations provide a mechanism for bringing others' knowledge and insights to bear on difficult-to-understand topics, allowing participants to capitalize on the collective knowledge of previous participants. Following from [38-40], who found that messages containing emotional affect also tend to be more cognitively engaging, our first hypotheses are that comments with emotional valence will *amplify* these deliberative effects:

H1a. Participants exposed to negative- and positive-valence annotations will feel more informed about their decision to consent or not than participants exposed to mixed-valence comments.

H1b. Participants exposed to negative- and positive-valence annotations will feel that they understand the content of the consent form better than participants exposed to mixed-valence comments.

Dinev and Hart [49] have asserted that concern for privacy issues is based on two processes: "(1) interaction with information technology (the Internet in this case), which requires a set of skills and a level of technical literacy, and (2) a social process of communication and transaction with sometimes anonymous or little-known social entities (companies or individuals) in the networked environment" (p. 8). Individuals with low levels of privacy concern therefore tend to have relatively basic mental models of privacy-related issues [50] and do not benefit from the predictive and explanatory power of informed mental models for understanding risky situations and interactions [50], as individuals who are knowledgeable about privacy-related issues do. Kittur et al [51] showed that having access to others' mental models and knowledge representations can help individuals build and refine their own schemas. Furthermore, considering Smith and Petty's [40] assertion that messages with negative affect can be more persuasive than those with positive affect, it is possible that these individuals with low privacy concern are also more prone to the influence of comments with negative valence. Thus:

H2a. The effect of exposure to social annotation on the extent to which participants feel informed will be stronger for individuals with lower privacy concern when exposed to negative valence comments than when exposed to mixed- or positive-valence comments.

H2b. The effect of exposure to social annotation on the extent to which participants feel they understand the content of the consent form will be stronger for individuals with lower privacy concern when exposed to negative-valence comments than mixed- or positive-valence comments.

We propose that annotation valence also plays a role in how individuals assess the trustworthiness of the organization seeking consent. Prior research has examined the role of technology-mediated social influence in protecting users in trust-related situations such as security and privacy threats [30], as well as from antisocial or exploitative behavior [52]. Potential concerns, shared by prior users, about the information provided to current users may therefore influence their perception of the information [53]. Prior studies demonstrate that negatively framed information is significantly more effective than positively framed information in shaping users' perceptions [54,55]. In the context of trust, researchers note a "negativity bias" for information communicating risk [56]. That is, individuals tend to trust negative messages more than they trust positive messages in the context of hazard-related information. We therefore propose the following:

H3a. Participants exposed to negative-valence annotations will trust the organization less than participants exposed to either mixed- or positive-valence comments.

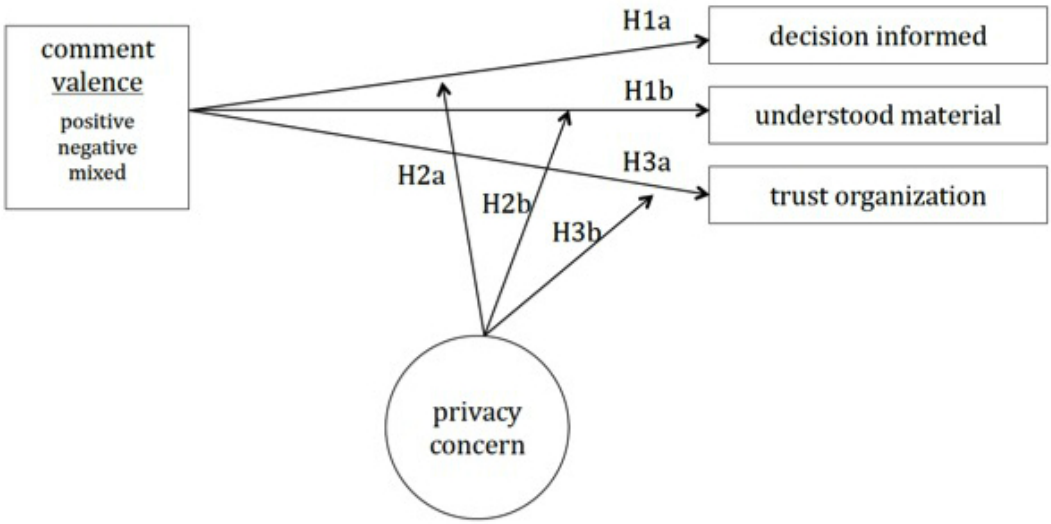
Prior research has shown that individuals with high and low privacy concern form trust in online contexts differently from each other [57] and that individuals' existing attitudes about a topic can moderate the effect of message valence on trust [58]. Specifically, Petty and Cacioppo [59] suggest that individuals with less experience in a topic are more likely to focus on peripheral cues (such as design or reputation) compared with experienced individuals to infer trustworthiness. Taken together with findings around the "negativity bias" associated with communicating risk [56], we propose the following:

H3b. The effect of exposure to social annotation on the extent to which participants trust the organization will be stronger for individuals with lower privacy concern when exposed to negative-valence comments than mixed- or positive-valence comments.



XSL•FO
RenderX

**Figure 1.** Research model depicting dependent, independent, interaction terms, and study hypotheses.




## Methods

### Procedure

We conducted a between-subjects experimental study to explore the effects of message valence in online social annotations on users' beliefs and behavior surrounding consent.

A website was developed specifically for this experiment. A link to the study was made available on Amazon Mechanical Turk, and participants were paid US $5.00 for completing the questionnaires. Participation in the study was limited to English speakers with a record of at least 100 prior tasks at an approval rate exceeding 99%. Since DTCGT is marketed to the general population, we chose to recruit users via Amazon Mechanical Turk. The population of Amazon Mechanical Turk is diverse and reflective of the general population, making it a viable venue for data collection [60,61].The choice of high prior approval rate and the relatively high pay was made in order to increase the likelihood that participants will be reliable and that they will take their time when considering the various choices they have to make as they go through the study.

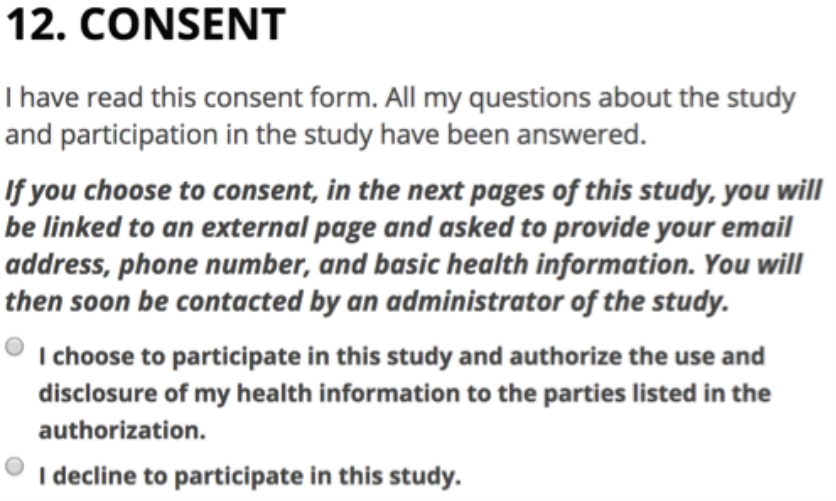
Participants were asked to take part in a study seeking to understand how users engage and learn from personal genomic information. They were first asked to answer several questions about their Internet usage (ie, privacy questionnaire) and to complete a tutorial on genomics. They were then asked to review the consent form for an *additional* study in which they could participate that would result in the mapping of their own genome. Users were randomly assigned to view an online consent form with social annotations that exhibited positive, negative, or mixed valence.

In order to maintain ecological validity, participants were led to believe that the additional genome mapping study was a real study in which they could participate. Participants were told that if they consented, they would be linked to an external page where they would be asked to provide their email address, phone number, and basic health information and would be contacted by an administrator of the genomics study to coordinate further (Figure 2). This deception was used to increase the likelihood that participants would take the time to make an informed and honest decision based on the information provided in the consent form. We did not disclose to participants that the genomic study was fictional until the end of the Mechanical Turk study when they were told the true objective of the study was to learn about the process of consent. No identifying information (email, phone number, etc) was ultimately collected.



XSL•FO
RenderX

**Figure 2.** Consent question used in study.




## Research Instruments

### *Privacy Questionnaires*

A privacy questionnaire and personal genomics tutorial preceded the consent form. Because the majority of the risks and issues with digitally mediated research center on data privacy, particularly in the context of genomics research, we used a measure of pre-existing privacy concern to assess an individual's existing attitude towards online privacy-related issues. We used a validated 16-item measure for privacy concern developed by Buchanan et al [62] based on Westin's privacy index [63] (see Table 1). Each question was answered using a 5-point Likert scale between "Not at all concerned" and "Extremely concerned."

**Table 1.** Buchanan et al's [62] measure of privacy concern.

| Question # | Question content |
|---|---|
| 1 | In general, how concerned are you about your privacy while using the Internet? |
| 2 | Are you concerned about online organizations not being who they claim they are? |
| 3 | Are you concerned that you are asked too much personal information when you register or make online purchases? |
| 4 | Are you concerned about online identity theft? |
| 5 | Are you concerned about people online not being who they say they are? |
| 6 | Are you concerned that information about you could be found on an old computer? |
| 7 | Are you concerned who might access your medical records electronically? |
| 8 | Are you concerned about people you do not know obtaining personal information about you from your online activities? |
| 9 | Are you concerned that if you use your credit card to buy something on the Internet your card number will be obtained/intercepted by someone else? |
| 10 | Are you concerned that if you use your credit card to buy something on the Internet your card will be mischarged? |
| 11 | Are you concerned that that an email you send may be read by someone else besides the person you sent it to? |
| 12 | Are you concerned that an email you send someone may be printed out in a place where others could see it? |
| 13 | Are you concerned that a computer virus could send out emails in your name? |
| 14 | Are you concerned about emails you receive not being from whom they say they are? |
| 15 | Are you concerned that an email containing a seemingly legitimate Internet address may be fraudulent? |

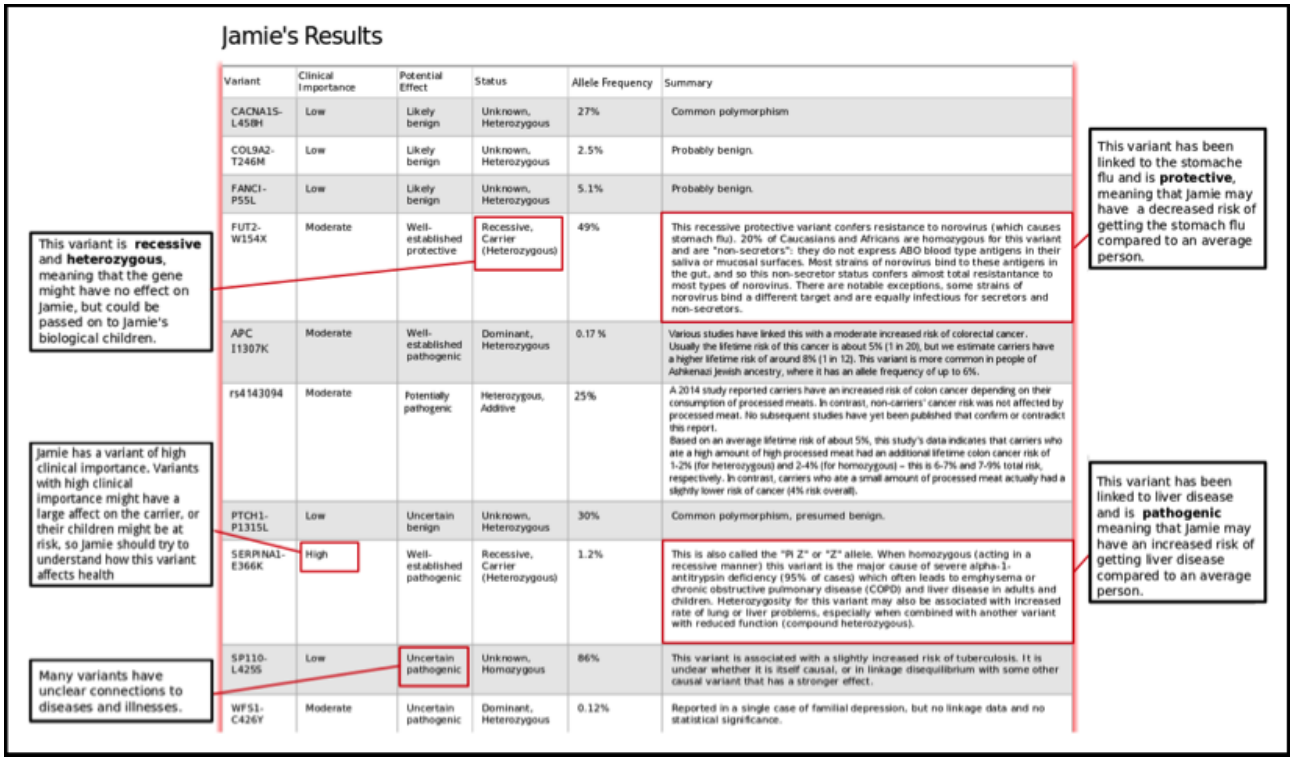
### *Genomics Tutorial*

The personal genomics tutorial comprised learning materials on the human genome and personal genomics developed by the Personal Genetics Education Project [64]. Participants' understanding of the material was assessed using a short 6-question quiz. Participants were then presented with a sample personal genomics report for an imaginary individual named Jamie, followed by another comprehension task. This task was used to demonstrate the type of information provided by genetic testing. Jamie's report was developed for this study using a fictional dataset in which sex and ethnicity did not have a specific effect and was modeled on GET-Evidence [64], Harvard's Personal Genomes Project's personal genomics report. Participants were asked to study the report and to answer three comprehension questions. Figure 3 shows the personal genomics report presented to users.



XSL•FO
RenderX

**Figure 3.** Sample genomic report presented to users in the training portion of this study.




| Variant | Clinical Importance | Potential Effect | Status | Allele Frequency | Summary |
|---|---|---|---|---|---|
| CACNA1S-L458H | Low | Likely benign | Unknown, Heterozygous | 27% | Common polymorphism |
| COL9A2-T246M | Low | Likely benign | Unknown, Heterozygous | 2.5% | Probably benign. |
| FANCI-P55L | Low | Likely benign | Unknown, Heterozygous | 5.1% | Probably benign. |
| FUT2-W154X | Moderate | Well-established protective | Recessive, Carrier (Heterozygous) | 49% | This recessive protective variant confers resistance to norovirus (which causes stomach flu). 20% of Caucasians and Africans are homozygous for this variant and are "non-secretors": they do not express ABO blood type antigens in their saliva or mucosal surfaces. Most strains of norovirus bind to these antigens in the gut, and so this non-secretor status confers almost total resistantance to most types of norovirus. There are notable exceptions, some strains of norovirus bind a different target and are equally infectious for secretors and non-secretors. |
| APC I1307K | Moderate | Well-established pathogenic | Dominant, Heterozygous | 0.17 % | Various studies have linked this with a moderate increased risk of colorectal cancer. Usually the lifetime risk of this cancer is about 5% (1 in 20), but we estimate carriers have a higher lifetime risk of around 8% (1 in 12). This variant is more common in people of Ashkenazi Jewish ancestry, where it has an allele frequency of up to 6%. |
| rs4143094 | Moderate | Potentially pathogenic | Heterozygous, Additive | 25% | A 2014 study reported carriers have an increased risk of colon cancer depending on their consumption of processed meats. In contrast, non-carriers' cancer risk was not affected by processed meat. No subsequent studies have yet been published that confirm or contradict this report. Based on an average lifetime risk of about 5%, this study's data indicates that carriers who ate a high amount of high processed meat had an additional lifetime colon cancer risk of 1-2% (for heterozygous) and 2-4% (for homozygous) – this is 6-7% and 7-9% total risk, respectively. In contrast, carriers who ate a small amount of processed meat actually had a slightly lower risk of cancer (4% risk overall). |
| PTCH1-P1315L | Low | Uncertain benign | Unknown, Heterozygous | 30% | Common polymorphism, presumed benign. |
| SERPINA1-E366K | High | Well-established pathogenic | Recessive, Carrier (Heterozygous) | 1.2% | This is also called the "Pi Z" or "Z" allele. When homozygous (acting in a recessive manner) this variant is the major cause of severe alpha-1-antitrypsin deficiency (95% of cases) which often leads to emphysema or chronic obstructive pulmonary disease (COPD) and liver disease in adults and children. Heterozygosity for this variant may also be associated with increased rate of lung or liver problems, especially when combined with another variant with reduced function (compound heterozygous). |
| SP110-L425S | Low | Uncertain pathogenic | Unknown, Homozygous | 86% | This variant is associated with a slightly increased risk of tuberculosis. It is unclear whether it is itself causal, or in linkage disequilibrium with some other causal variant that has a stronger effect. |
| WFS1-C426Y | Moderate | Uncertain pathogenic | Dominant, Heterozygous | 0.12% | Reported in a single case of familial depression, but no linkage data and no statistical significance. |

### *Social Consent Form*

Following the genomics tutorial, participants were presented with the consent form for an additional, optional study in which their genomes would be mapped and their family health history and trait information would be collected online. The study was framed as a voluntary contribution to research (rather than a commercial service in exchange for payment), but those who chose to participate would receive their results in a free, online report. The content of the consent form was based on Office for Human Research Protections guidelines [65], the Personal Genome Project consent form [66], and the 23andMe informed consent document (publicly available online [67]). Modifications to improve the clarity of the text were made based on feedback provided in pilot tests with other Amazon Mechanical Turk users.

The experimental consent form included comment boxes with social annotations in the margins of the screen (Figure 4). Participants were told that these annotations had been contributed by previous prospective participants who had seen the same consent form. In reality, the content was derived from feedback provided by participants during earlier pilot tests and included questions, concerns, personal perspectives, and contextual information related to the content of the consent form. We used our best judgment to select feedback in which the sentiment expressed was not unreasonably extreme. The selected comments were then edited such that each had positive and negative valence versions of itself, allowing us to standardize and control the topics of social annotations across conditions. Though they were manipulated, deriving the annotations from real content allowed us to use material that touched on topics likely to be meaningful to current participants.

The three experimental conditions included one iteration of the consent form in which the onscreen annotations contained all of the positive-valence comments, one iteration that contained only the negative-valence comments, and a final iteration that contained mixed-valence comments: positive and negative valence comments were alternated equally in the text, beginning with a positive-valence comment. To compare across these conditions, we placed comments at the same point in the text, referencing the same passages and topics in the text of the consent form.

Prior research on the effects of message valence has largely compared positive- to negative- valence messages to each other, or messages containing some valence with neutral messages. Participants' feedback in early stages of the study indicated that comments in this context are rarely neutral: personal genomics is an important topic that evokes emotionally charged responses. To preserve ecological validity, we therefore chose to examine the effects of mixed-valence annotations rather than neutral annotations or annotations whose overall effect was neutral.

Annotations in each condition also displayed an indicator showing how many other (hypothetical) study participants "liked" the comments. The number of "likes" for each comment was determined by the researchers and ranged from 0-46 likes on a comment. The same number of likes were displayed for each comment, in each condition (ie, both the positive and negative valence instances of a comment in each of the three conditions had the same number of likes).

Participants in this study had the ability to interact with the annotations and likes embedded in the consent form (unlike in our first study where the comments were entirely static). We wanted to provide the participants the opportunity to engage with the annotations more directly and in ways that you might



XSL•FO
RenderX

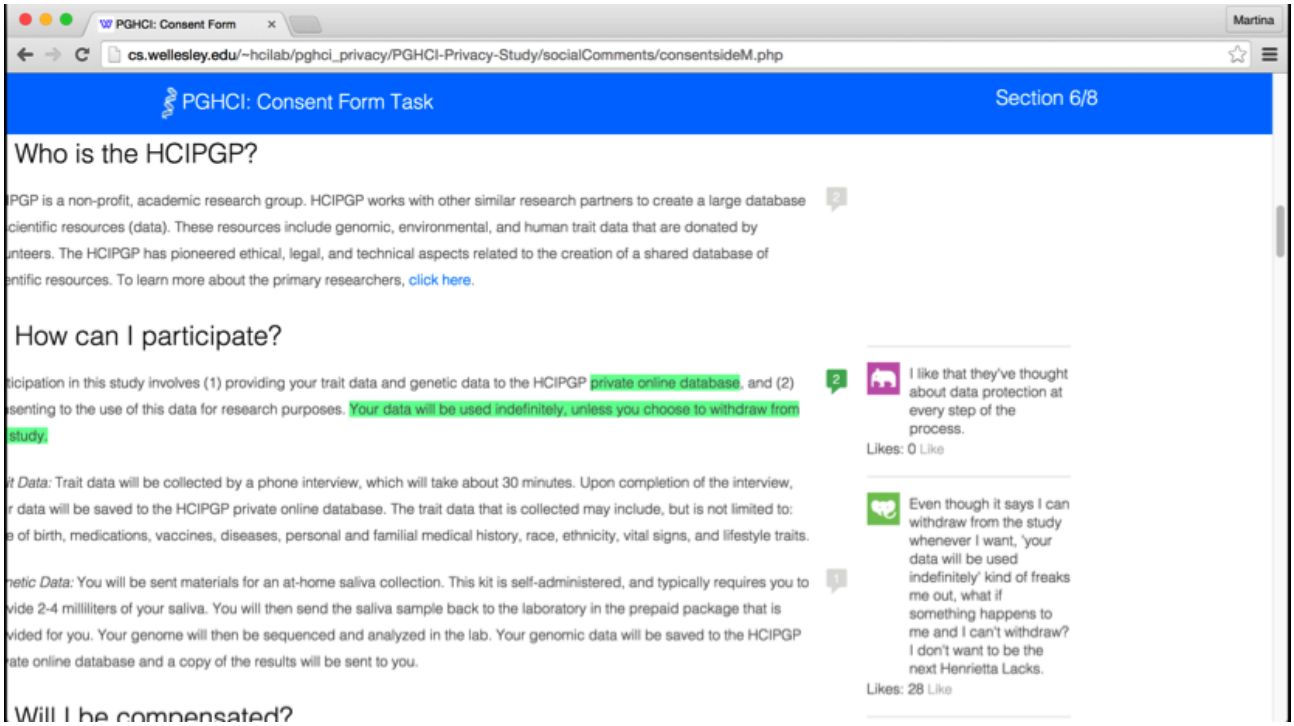
find elsewhere online. In our study, we used the SideComments application programming interface to implement functionality that allowed participants to respond to or "like" existing comments or to create their own highlights and textual annotations. They could also click on a comment to open or close it or could hover over an in-text highlight to open the associated comment. Stylized profile photos were used to improve the ecological validity of the annotations: websites that incorporate social annotations frequently implement some mechanism for signaling to participants that the comments came from multiple authors.

To ensure that the added level of interactivity did not present a confound in our study of message content, we devised and tested an iteration of the interface in which the comments were non-interactive. The comments were identical in message and placement to the annotations in the interactive mixed-valence condition. We recruited 137 participants and presented them with the same study as participants in the interactive conditions, and Student's *t* tests were used to compare measure ratings between the interactive and non-interactive conditions with mixed-valence comments. The differences between the two conditions were not statistically significant in any of the measures examined in this study (see Table 2). We can therefore conclude that the additional level of interactivity does not present a confound in our study.

**Table 2.** Comparison of measures between an interactive, mixed-valence condition, and a non-interactive, mixed-valence condition.

| | Interactive, mixed-valence condition | Non-interactive, mixed-valence condition | *P* |
|---|---|---|---|
| **Perceptual, mean (SD)** | | | |
| Decision was informed | 4.5 (0.69) | 4.46 (0.65) | .74 |
| Understood all the material | 4.19 (0.93) | 4.25 (0.76) | .72 |
| Trust the organization seeking my consent | 3.82 (0.82) | 3.66 (0.94) | .26 |
| **Consent** | | | |
| Consent, n (%) | 20 (43%) | 65 (49%) | .61 |
| No consent, n | 26 | 67 | |

**Figure 4.** Screenshot of consent form with highlighted text and social annotations.




## Measures

Following their decision to consent to the personal genomic study described in the consent form, users were presented with questions about their deliberative process and perceptions of the consent form (see Table 2). All measures were single-item and self-reported using a 5-point Likert scale (strongly disagree to strongly agree). Studies on informed consent have traditionally equated how informed a participant felt with how well they understood the material and therefore used comprehension tests of subject matter to infer informed consent [68] or to assess participants' ability to give informed consent [69]. Similarly, trust was historically measured using trust games [70] to form an "objective" measurement. In this study, we were interested in the *perception* of feeling informed, of having



XSL•FO
RenderX

understood the material, and that the organization is trustworthy. We therefore drew from Sepucha et al's [71] single-item measure of the perception of being informed. The measures for their perception of understanding and trust are modifications of that question and contextually relevant. Table 3 lists the questions used to address each hypothesis.

**Table 3.** Questions used to evaluate each hypothesis.

| Hypothesis | Question |
|---|---|
| H1a, H2a | I feel that my decision (to consent or not) was an informed decision. |
| H1b, H2b | I feel that I understood the material presented and I have no additional questions. |
| H3a, H3b | Based on what I have seen and read in this consent form, I feel like I can trust the HCIPGP to use and protect my data in the ways outlined in the consent form. |

### Demographics and Disclosure

Prior research has shown that demographic variables can influence how informed participants feel [71] and an individual's likelihood of participation in medical research [72]. We therefore collected demographic data that included education, age, and gender. After answering the demographic questions, they were informed that the study was fictitious and that the true research question related to the process of consent and consent forms.

### Data Analysis

Analysis of variance with covariates was used to identify main effects of condition and interaction effects where applicable, while controlling for demographic variables and participants' pre-existing attitude towards information privacy. Post-hoc Tukey tests were performed to further examine the results pairwise. The interactivity measures (ie, number of times participants opened, liked, or hover over comments, and how many comments they wrote) were found to contain positive skew (ie, a larger number of participants interacted relatively little with the interactive features of the consent form). To correct for this skew and produce a relatively symmetrical distribution of actions, we transformed the counts for each interactive measure by using its square root in the analysis [73].

## Results

### Demographics

A total of 152 participants took part in this study: 56 participants were assigned to the negative valence condition, 46 participants to the mixed valence condition, and 47 participants to the positive valence condition. The average age of participants was 34.25 years (SD 10.78), and 72 (48.3%) participants were female. One participant had some high-school education, 12 participants had high school diplomas, 58 participants had some college education, 59 participants had bachelor degrees, 14 participants had master's degrees, 3 participants had doctoral degrees, and 2 participants declined to state their education.

### Domain Comprehension

Participants spent 3.88 minutes on average (SD 3.14 min) studying the genomics tutorial, and 3.96 minutes on average (SD 2.21 min) studying Jamie's sample genomics test results. Only 3 (out of 152) answered fewer than 3 out of 6 genome tutorial questions, or fewer than 2 out of 3 of the genome report questions, incorrectly. These individuals were removed from the dataset, leaving 149 viable participants.

Correlation analysis was used to test whether the domain comprehension scores from the entire population impacted the extent to which they felt their decision was informed (ie, informed consent). Within the subset of viable participants, the correlation analysis between participants' comprehension scores and perceptual variables failed to reach significance. The domain comprehension score was therefore not controlled for going forward.

Participants had a mean rating of 2.93 (between 1 and 5, SD 0.87) on our measure of privacy concern.

### Time on Consent Form

In the condition with the negative-valence comments, participants spent an average of 7.57 minutes (SD 8.56 min) studying the consent form before deciding whether to consent. In the mixed condition, participants spent 8.18 minutes (SD 7.14 min), and in the positive condition participants spent 5.82 min (SD 4.20 min) prior to deciding whether to consent. An analysis of variance testing the distribution of time across conditions shows that condition does not a have a significant main effect on time: the amount of time spent studying the consent form did not differ significantly between social annotations' valence. We did observe, however, a significant effect of gender on time: female participants took significantly longer to read the consent form (mean 488.72, SD 488.04) than male participants (mean 357.38, SD 300.65; $F_1$=6.177, $P$=.014).

Overall, participants who consented spent significantly less time studying the consent form than participants who did not consent (mean 5.62 min, SD 7.36 min and mean 8.39, SD 5.86 min, respectively; $F_1$=6.477, $P$=.012). Further inspection shows that the difference in time to consent differed significantly only in the positive affect condition: participants who consented spent significantly less time (mean 4.05 min, SD 2.77 min) studying the consent form than participants who did not consent (mean 7.88 min, SD 4.18 min; $F_2$= 14.3, $P$<.001). The time spent in the other three conditions did not differ significantly between those who did and did not consent.

### Interactivity Measures

The number of times participants liked, opened, or added comments to the consent form did not differ significantly across conditions (see Table 4). The number of times participants hovered over in-text highlights, however, did differ significantly



XSL•FO
RenderX

by condition (*P*=.008). Specifically, participants in the mixed-valence condition (mean 4.36, SD 6.50) were significantly more likely to hover over highlights than participants in the positive condition (mean 1.56, SD 2.98). We also found a marginally significant effect of age on behavior: older participants tended to hover over the in-text highlights marginally more frequently than younger participants (*F*=2.86, *P*=.09). The differences between the negative and mixed conditions, and the positive and negative conditions, on the other hand, failed to reach significance.

## Dependent Variables

Our main findings are presented in Table 4.

**Table 4.** Results from the comparison between the negative-, mixed-, and positive-valence conditions.

| | Negative valence comments | Mixed valence comments | Positive valence comments | *P* |
|---|---|---|---|---|
| **Perceptions, mean (SD)** | | | | |
| Decision was informed | 4.45 (0.63) | 4.5 (0.69) | 4.17 (0.94) | .07 |
| Understood all the material | 3.98 (1.05) | 4.19 (0.92) | 4.28 (0.69) | ns |
| Trust the organization seeking my consent | 3.59 (1.14) | 3.82 (0.82) | 4.02 (0.90) | .08 |
| **Interaction terms, mean (SD)** | | | | |
| Liked comments | 1.43 (2.62) | 1.80 (2.52) | 1.53 (2.67) | ns |
| Commented | 1.62 (3.04) | 1.61 (2.27) | 1.19 (1.65) | ns |
| Opened comment | 5.46 (7.30) | 7.54 (9.11) | 5.72 (5.87) | ns |
| Hovered over in-text highlight | 2.88 (7.61) | 4.36 (6.50) | 1.56 (2.98) | .08 |
| Time (s) | 454.12 (513.87) | 461.89 (392.68) | 341.00 (234.09) | .012 |
| **Behavior** | | | | |
| Consent, n (%) | 27 (48.21%) | 20 (43.48%) | 27 (57.44%) | ns |
| No consent, n | 29 | 26 | 20 | |

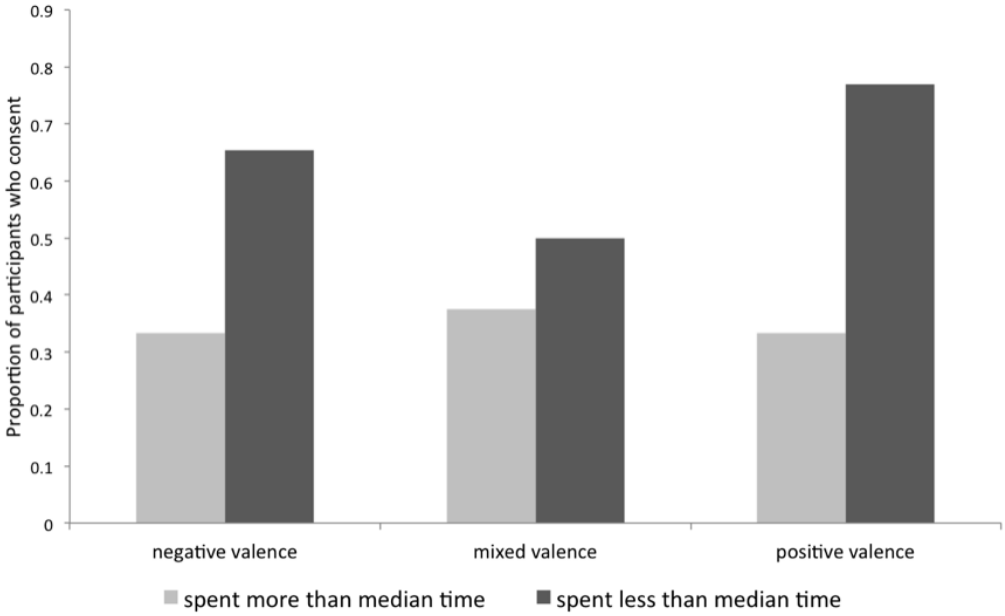
## Consent

The rate of consent did not differ significantly across conditions: 48% (27/56) of participants consented in the negative valence condition, 43% (20/46) consented in the mixed-valence condition, and 57% (27/47) consented in the positive condition. There was, however, a significant interaction between condition and the amount of time participants spent studying the consent form on the consent rate (*Z*=-2.686, *P*=.007): participants in the negative condition were less likely to consent when they spent more time studying the consent form. Participants exposed to the positive condition, on the other hand, were significantly more likely to consent than participants in other conditions when they had spent less time studying the consent form (Figure 5). We also found a marginally significant effect of age on the probability of consenting to the study: the mean age of participants who consented (mean 32.63, SD 8.95) was marginally lower than participants who did not consent (mean 35.85, SD 12.17; $F_1$=3.57, *P*=.06).



XSL•FO
RenderX

**Figure 5.** Proportion of participants who consented in each condition depending on whether they spent more or less than the median amount of time studying the consent form.




## Perceptions About Consent

### *Decision Was Informed*

The experimental intervention had a marginally significant main effect on participants' beliefs ($F_2$=2.40, $P$=.07). Specifically, Tukey's post-hoc tests indicate that participants in the positive condition (mean 4.17, SD 0.94) felt marginally less informed than those in the mixed condition (mean 4.50, SD 0.69, $P$=.09), though the differences between ratings in the mixed- and negative-valence conditions and the positive and negative conditions were not significant. We therefore reject hypothesis H1a. We also reject hypothesis H2a as there appear to be no significant interactions between condition and privacy concern, or a main effect of privacy concern on how informed a participant felt. We did find, however, a significant main effect of gender on the dependent variable. Specifically, female participants felt more informed (mean 4.5, SD 0.65) than male participants (mean 4.26, SD 0.85; $F_1$=5.151, $P$=.02).

### *Understood the Material*

Our results indicate that condition does not have a main effect on participants' belief that they understood the content of the consent form, and this effect does not differ according to participants' prior privacy preserving attitudes and behavior. We therefore reject hypotheses H2a and H2b.

### *Trust the Research Organization*

Condition had a marginal main effect on the extent to which participants reported trusting the organization ($F_2$=2.566, $P$=.08). In particular, Tukey post-hoc tests show that participants in the negative condition (mean 3.59, SD 1.14) were marginally less trusting than participants exposed to the positive condition (mean 4.02, SD 0.90, $P$=.06) in partial support of hypothesis H3a. However, neither participants in the positive nor the negative conditions differed significantly from participants in the mixed-valence condition.

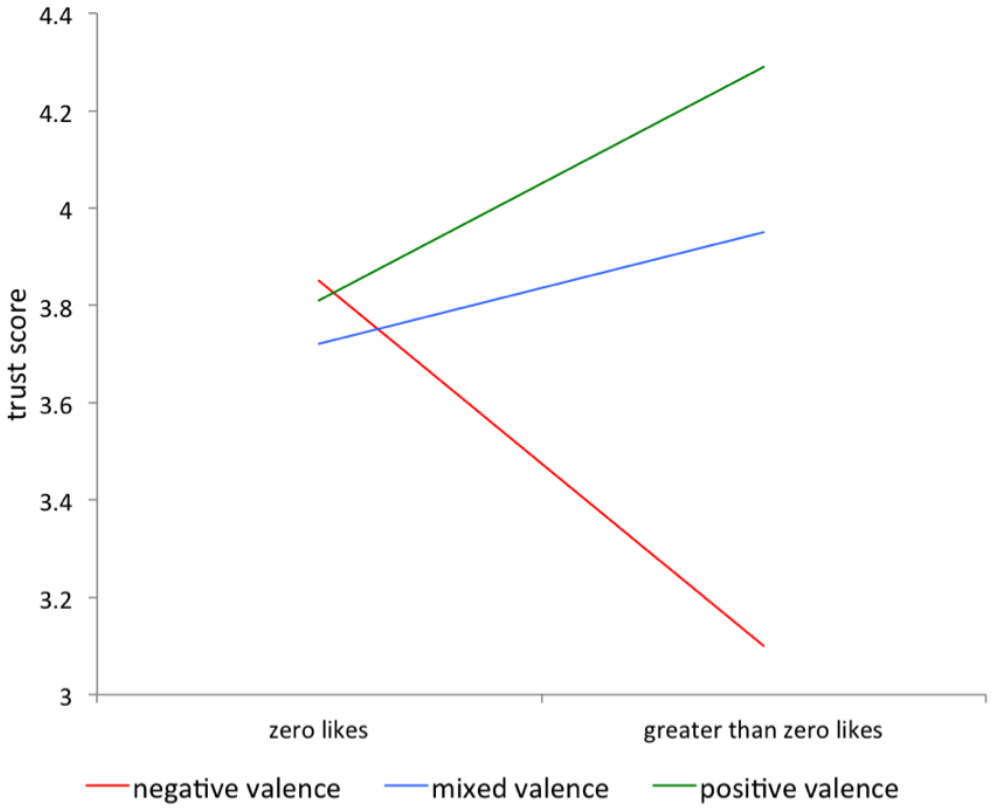
Although we observed a significant, negative main effect of privacy concern on participants' trust in the organization ($F_1$=12.80, $P$=.0005), the interaction between condition and participants' privacy concern failed to reach significance, leading us to reject hypothesis H3b. We do, however, find a significant interaction between the experimental intervention and the number of times participants clicked "like" next to an annotation ($F_1$=3.47, $P$=.04): participants who clicked "like" a high number of times reported trusting the organization less when exposed to the negative condition than participants exposed to the mixed- and positive-valence conditions. Because a high proportion of participants never clicked "like" (59%, 88/149 of participants), Figure 6 depicts this interaction based on whether or not the user use "like" button.

We also observed a marginally significant effect of age (older participants tended to trust the organization less than younger participants: B=-0.01, $t_{148}$=-1.89, $P$=.06) and gender (male participants tended to trust the organization more [mean 3.93, SD 0.93] than female participants [mean 3.65, SD 0.94]; $t_{148}$=1.692, $P$=.09) on the extent to which a participant trusted the organization.



XSL•FO
RenderX

**Figure 6.** Impact of the interaction of condition and number of likes on the extent to which participants reported trusting the organization.




## Discussion

### Principal Findings

In this study, we found that the valence communicated in social annotations, which are embedded in an interactive informed consent form, can influence individuals' perceptions and beliefs about consent. In particular, we show that consent forms containing positive valence annotations are likely to lead participants to feel less informed and simultaneously more trusting of the organization seeking consent. In certain cases where participants spent little time considering the content of the consent form, participants exposed to positive valence annotations were even more likely to consent to the study.

While our findings that participants in the mixed-valence condition felt more informed than participants in the positive-valence condition may seem surprising in the context of previous studies comparing positive- and negative-valence messages, we argue that it contributes to our understanding of social influence in contexts where sentiment is effectively mixed. Prior research shows that individuals tend to focus on the negative elements of the consent process as a result of the information provider's desire to warn others about threats, and the information seeker's desire to acquire more information about a potential problem [30]. Drawing attention to limitations of the consent form using social annotation highlights the limitations in participants' own knowledge and the shortcomings of the consent form, contributing to participants' simultaneously feeling more informed [6]. By this logic, participants in the positive condition would feel less informed because these limitations would be trivialized or framed positively, which is indeed consistent with the relatively lower ratings of feeling informed measured in this study. Participants in the mixed condition may report feeling relatively more informed precisely because the comments are both positive and negative: alternating valence may engage participants around negative aspects of the consent form as well as create the perception of debate and deliberation with the addition of positive comments. Participants who have mixed feelings or are conflicted around issues presented in the consent form may be able to match their needs more easily and engage more deeply with variegated valence [37]. Another explanation may follow from Smith and Petty [40] who found that messages with unexpected framing, regardless of valence, tend to be more cognitively engaging. It is possible that comments with mixed valence are more "surprising" to participants and therefore more engaging than instances where they can expect that the annotations will be positively or negatively framed. Regardless, existing research on mixed-valence social annotations is sparse: authors focus on comparing positive to negative valence comments [38] or comments containing valence to neutral comments [42]. This study therefore contributes to the relatively understudied (and more ecologically valid) instances where valence is mixed.

Our results show that participants' trust in the organization also differs across condition: participants in the negative valence condition were significantly less trusting than participants in the positive valence condition. This finding is supported by previous research showing that negative messages tend to be more persuasive in general [40]. From the literature on consent processes, we also know that people tend to look to socially constructed information to understand the negative aspects of



XSL•FO
RenderX

consenting (eg, risks and consequences) rather than the positive aspects (eg, the benefits of participation) [30]. This interpretation is supported by the significant interaction observed between the valence condition and the number of times participants clicked "like." "Liking" a comment is an explicit way for an individual to agree with the questions, perspectives, or opinions of the author of the comment. In this study, we observed that participants who agreed more frequently with negative valence comments reported trusting the organization less than participants who agreed with comments less frequently, or participants in other conditions. We argue that this highlights the persuasive nature of these types of comments.

Notably, even when valence was extreme (as in the positive and negative manipulations), there was no significant impact on the ultimate metric of consent rates. This seems to indicate that implementing social comments on consent processes may risk little in terms of actual consent rates, while giving participants an increased sense of autonomy by helping them feel more informed. This is generally consistent with the results of our previous study [6], in which we found that consent rate did not differ significantly between a condition containing social annotations and a control. The interaction effect between time and condition is a surprising and important result, however, because it calls into question the tenet of voluntariness for informed consent for participants exposed to comments with positive valence: participants in that condition who studied the consent form for less time were more likely to consent. This result may be explained by Joyce and Kraut's [45] findings that messages containing positive valence tend to evoke a sense of community and encourage individuals to participate, whereas messages with negative valence provoke heated exchanges. It may be the case that participants who spend less time considering the content of the consent form are more susceptible to these effects, whereas participants who spend more time engaging with the material and debating the content on their own are more likely to act on their own opinions of the content.

## Contributions

This study has demonstrated that social annotation interventions can have an impact in a biomedical informed consent decision-making context. In contrast to the spaces where social annotation studies have traditionally been conducted (eg, consumer products, online search platforms, and security feature adoption), human subjects research requires decisions that are intensely personal and can have substantial ramifications for the individual as well as their families. Our research demonstrates that *strangers'* perspectives, knowledge, and opinions can play a significant role in how individuals make these decisions for themselves, implying a shift in the way that we think about and execute consent-seeking processes.

Social influence in online environments and its effect on users in social recommender systems has been the topic of substantial research in recent years [30,74,75]. These studies have largely examined the effects of explicit organizational and social structures (eg, interpersonal relationships, professional hierarchy, physical proximity) on social influence [27]. Our study contributes to this body of literature by exploring the impact of anonymous message content, and in particular, the emotional valence communicated in messages, on social influence in socially enabled, digitally mediated consent processes when explicit organizational and social structures are necessarily missing due to the sensitive context of biomedical research.

Our results also contribute to the literature on valence in social annotation. The existing research on mixed-valence social annotations is sparse: authors focus on comparing positive to negative valence comments [38] or comments containing valence to neutral comments [42]. It is rarely the case that the annotations in a document will be uniformly negative or positive; this study therefore contributes to our understanding of the relatively understudied, yet frequent, instances where valence is mixed.

This study represents a new and expanded understanding of the multidimensionality of social annotation in a high-risk decision-making context. Our previous study showed that the inclusion of social annotation does not merely improve or worsen the user's experience (as put forth in existing studies); rather, it changes how participants reflect on their ability to make informed decisions for themselves in complex ways. Here we extend that line of research to provide a unique and nuanced perspective on how inherent qualities of user-generated content, namely emotional valence, can influence and engage individuals. This is particularly salient in the context of informed consent because the focus of deliberation is not among members for the purpose of consensus agreement, but within the individual [76]. These findings may be further expanded to inform the decisions around how comments are to be implemented. The designers of systems containing user-generated content must decide whether or not to moderate user-generated comments—a decision for which we have outlined several important considerations with this research.

## Limitations and Future Research

While this study demonstrates how exposure to computer-supported social annotations impacts individuals' perceptions in the context of informed consent, it has a number of limitations. Though we believe that the demonstrated increase in the perception of being informed suggests that social annotations can benefit prospective participants, the experiment was structured to study the effects of exposure to annotations on participants' *perceptions* and did not examine whether they objectively benefitted from the intervention. Future research is needed to explore whether improvements in the perception of making an informed decision we observed result in quantifiable and objective improvements in the process of analyzing complex consent forms, and whether it results in objectively "better" outcomes for the individual.

Furthermore, we look at the impact of a narrow range of emotional valence that is operationalized in their extremes; that is to say that it is unlikely that the user will be confronted with only positive, only negative, or perfectly mixed-valence comments. It is more likely that they would be confronted with some complex mix of the two that leans toward an overall positive or negative effect. Furthermore, we prioritized using



XSL•FO
RenderX

ecologically representative comments in our study rather than controlling for the strength of sentiment contained in each comment individually. Additional research is needed to control for and understand the impact of less extreme and less consistent examples of emotional valence [77].

The small sample sizes used in this study may also have obscured findings related to participants' perceptions, given that the manipulation of sentiment was relatively subtle. We believe that the results we have presented here are compelling for an exploratory study such as this one, but future research should consider larger sample sizes when investigating related questions.

Finally, a number of important questions remain for further investigation that will help us determine whether social annotation interventions are appropriate in this context. Evaluating the effect of creating and actively engaging with social annotation on user behavior requires us to understand how to solicit meaningful content from participants, what motivates individuals to contribute content, what privacy issues are associated with contributing and accessing health-related information, and how (or whether) to "police" information contributed by anonymous others in a form with such a significant impact: additional research is needed to understand whether moderating user-contributed information to create the desired effect is ethical and effective. Knowing that we may be able to improve certain aspects of the process of deliberating consent by incorporating novel and non-traditional sources of information, however, obligates us as a community to explore social annotation interventions further.

## Conclusion

Electronic consent has become increasingly popular in Internet research in general and biomedical research in particular. The work presented here explores the effects of adding a computer-supported social dimension, which inherently contains human emotions and opinions, to the consent deliberation process. In our first study we found that exposure to social annotations results in participants' feeling that their decision was more informed, but simultaneously less confident in their understanding of the genomics material presented in the consent form as well as less trusting of the organization soliciting the consent. Based on these findings, we proposed that augmenting the consent deliberation process with multiple voices can enable individuals to capitalize on the knowledge of others, which brings to light questions, problems, and concerns they may not have considered on their own. In this study, we examined the influence of human emotion contained in these voices on participants' perceptions and beliefs about consent. We found that consent forms containing positive valence annotations are likely to lead participants to feel less informed and simultaneously more trusting of the organization seeking consent. In certain cases where participants spent little time considering the content of the consent form, participants exposed to positive valence annotations were even more likely to consent to the study. We suggest that these findings represent important considerations for the designers of such systems. We also call for future research that may extend the research on socially enabled online consent forms to examine the role of novel user-generated sources of information, and may develop new measures and indicators for evaluating social informed consent.

## Acknowledgments

This work was partially funded by Grants IIS-1017693 and IIS-1422706 from the National Science Foundation (NSF), Division of Information and Intelligent Systems (IIS).

## Conflicts of Interest

None declared.

## References

1. Kraut R, Olson J, Banaji M, Bruckman A, Cohen J, Couper M. Psychological Research Online: Report of Board of Scientific Affairs' Advisory Group on the Conduct of Research on the Internet. American Psychologist 2004;59:105. [doi: 10.1037/0003-066X.59.2.105] [Medline: 14992637]
2. Apple. ResearchKit. URL: http://www.apple.com/researchkit/ [accessed 2016-07-06] [WebCite Cache ID 6inU0zLaM]
3. Friedman B, Felten E, Millett L. Informed consent online: A conceptual model and design principles. In: University of Washington Computer Science & Engineering Technical Report. Seattle, WA: University of Washington; 2000.
4. Das S, Kramer A, Dabbish L, Hong J. The Role of Social Influence in Security Feature Adoption. 2015 Presented at: The 18th ACM Conference on Computer Supported Cooperative Work & Social Computing; 2015; Vancouver, Canada p. 1416-1426.
5. Center for Drug Evaluation and Research. Guidance Webinars. 2016. FDA Webinar: Draft Guidance: Use of an Electronic Informed Consent in Clinical Investigations Questions and Answers URL: http://www.fda.gov/Training/GuidanceWebinars/ucm440720.htm [accessed 2016-07-06] [WebCite Cache ID 6inVLKoMb]
6. Balestra M, Shaer O, Okerlund J, Ball M, Nov O. The Effect of Exposure to Social Annotation on Online Informed Consent Beliefs and Behavior. 2016 Presented at: The 19th ACM conference on Computer-Supported Cooperative Work and Social Computing; 2016; San Francisco, CA. [doi: 10.1145/2818048.2820012]
7. Forgas J, editor. Affect in Social Thinking and Behavior. United Kingdom: Psychology Press; 2012.
8. Shaer O, Nov O. HCI for personal genomics. Interactions 2014 Sep 01;21(5):32-37. [doi: 10.1145/2656622]



XSL•FO
RenderX

9. Shaer O, Nov O, Okerlund J, Balestra M, Stowell E, Ascher L, et al. Informing the Design of Direct-to-Consumer Interactive Personal Genomics Reports. J Med Internet Res 2015;17(6):e146 [FREE Full text] [doi: 10.2196/jmir.4415] [Medline: 26070951]
10. Kneipp S, Lutz B, Means D. Reasons for enrollment, the informed consent process, and trust among low-income women participating in a community-based participatory research study. Public Health Nurs 2009;26(4):362-369. [doi: 10.1111/j.1525-1446.2009.00791.x] [Medline: 19573215]
11. Millett L, Batya F, Edward F. Cookies and Web browser design: toward realizing informed consent online. 2001 Presented at: SIGCHI Conference on Human Factors in Computing Systems; 2001; Seattle, WA p. 46-52. [doi: 10.1145/365024.365034]
12. Culver C. Philosophy in medicine: conceptual and ethical issues in medicine and psychiatry. New York: Oxford University Press; 1982.
13. Ubel P, Loewenstein G. The role of decision analysis in informed consent: Choosing between intuition and systematicity. Social Science & Medicine 1997 Mar;44(5):647-656. [doi: 10.1016/S0277-9536(96)00217-1]
14. Gray B, Cooke R, Tannenbaum A. Research involving human subjects. Science 1978 Sep 22;201(4361):1094-1101. [doi: 10.1126/science.356268]
15. Morrow GR. How Readable Are Subject Consent Forms? JAMA 1980 Jul 04;244(1):56. [doi: 10.1001/jama.1980.03310010042026]
16. Duffy T, Kabance P. Testing a readable writing approach to text revision. Journal of Educational Psychology 1982;74(5):733-748. [doi: 10.1037/0022-0663.74.5.733]
17. Dresden G, Levitt M. Modifying a Standard Industry Clinical Trial Consent Form Improves Patient Information Retention as Part of the Informed Consent Process. Acad Emergency Med 2001 Mar;8(3):246-252. [doi: 10.1111/j.1553-2712.2001.tb01300.x]
18. Stiles P, Poythress N, Hall A, Falkenbach D, Williams R. Improving Understanding of Research Consent Disclosures Among Persons With Mental Illness. Psychiatric Services 2001 Jun;52(6):780-785. [doi: 10.1176/appi.ps.52.6.780]
19. Murphy D, O'Keefe Z, Kaufman A. Improving comprehension and recall of information for an HIV vaccine trial among women at risk for HIV: Reading level simplification and inclusion. AIDS Education and Prevention 1999;11(5):389.
20. Dunn L, Lindamer L, Palmer B, Golshan S, Schneiderman L, Jeste D. Improving Understanding of Research Consent in Middle-Aged and Elderly Patients With Psychotic Disorders. The American Journal of Geriatric Psychiatry 2002 Mar;10(2):142-150. [doi: 10.1097/00019442-200203000-00005]
21. Llewellyn-Thomas H, Thiel E, Sem F, Harrison Woermke D. Presenting clinical trial information: a comparison of methods. Patient Education and Counseling 1995 May;25(2):97-107. [doi: 10.1016/0738-3991(94)00705-Q]
22. Agre P, Rapkin B. Improving Informed Consent: A Comparison of Four Consent Tools. IRB: Ethics and Human Research 2003 Nov;25(6):1. [doi: 10.2307/3564285]
23. Fureman I, Meyers K, McLellan A, Metzger D, Woody G. Evaluation of a video-supplement to informed consent: injection drug users and preventive HIV vaccine efficacy trials. AIDS Educ Prev 1997 Aug;9(4):330-341. [Medline: 9376207]
24. Cattuto C, Barrat A, Baldassarri A, Schehr G, Loreto V. Collective dynamics of social annotation. In: Proc Natl Acad Sci. 2009 Jun 30 Presented at: Proceedings of the National Academy of Sciences 106; 2009; Philadelphia. PA p. 10511-10515 URL: http://www.pnas.org/cgi/pmidlookup?view=long&pmid=19506244 [doi: 10.1073/pnas.0901136106]
25. Gao F. A case study of using a social annotation tool to support collaboratively learning. The Internet and Higher Education 2013 Apr;17:76-83. [doi: 10.1016/j.iheduc.2012.11.002]
26. Nelson L, Held C, Pirolli P, Hong L, Schiano D, Chi E. With a little help from my friends: examining the impact of social annotations in sensemaking tasks. 2009 Presented at: SIGCHI Conference on Human Factors in Computing Systems; 2009; Boston, MA p. 1795-1798.
27. Cross R, Rice RE, Parker A. Information seeking in social context: Structural influences and receipt of information benefits. IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews) 2001;31(4):438-448. [doi: 10.1109/5326.983927]
28. Hu N, Liu L, Zhang J. Do online reviews affect product sales? The role of reviewer characteristics and temporal effects. Inf Technol Manage 2008 Mar 11;9(3):201-214. [doi: 10.1007/s10799-008-0041-2]
29. Ye Q, Law R, Gu B, Chen W. The influence of user-generated content on traveler behavior: An empirical investigation on the effects of e-word-of-mouth to hotel online bookings. Computers in Human Behavior 2011 Mar;27(2):634-639. [doi: 10.1016/j.chb.2010.04.014]
30. Das S, Kim T, Dabbish L, Hong J. The Effect of Social Influence on Security Sensitivity. 2014 Presented at: Tenth Symposium on Usable Privacy and Security; 2014; Menlo Park, CA p. 143-157.
31. Bernstien M, Tan D, Smith G, Czerwinski M, Horvitz E. Collabio: a game for annotating people within social networks. 2009 Presented at: 22nd Snnual ACM Symposium on User Interface Software and Technology; 2009; Victoria, Canada p. 97-100.
32. Evans B, Kairam S, Pirolli P. Do your friends make you smarter?: An analysis of social strategies in online information seeking. Information Processing & Management 2010 Nov;46(6):679-692. [doi: 10.1016/j.ipm.2009.12.001]
33. Kulkarni C, Chi E. All the news that's fit to read: a study of social annotations for news reading. 2013 Presented at: SIGCHI Conference on Human Factors in Computing Systems; 2013; Paris, France p. 2407-2416.




XSL•FO
RenderX

34. Cialdini R, Goldstein N. Social influence: compliance and conformity. Annu Rev Psychol 2004 Feb;55:591-621. [doi: 10.1146/annurev.psych.55.090902.142015] [Medline: 14744228]
35. Stieglitz S, Dang-Xuan L. Emotions and Information Diffusion in Social Media—Sentiment of Microblogs and Sharing Behavior. Journal of Management Information Systems 2013 Jun 7;29(4):217-248. [doi: 10.2753/MIS0742-1222290408]
36. Pantel P, Gamon A, Alonso G, Haas K. Social annotations: utility and prediction modeling. 2012 Presented at: 35th international ACM SIGIR Conference on Research and Development in Information Retrieval; 2012; Portland, OR p. 285-294. [doi: 10.1145/2348283.2348324]
37. Chen Y, Xie J. Online Consumer Review: Word-of-Mouth as a New Element of Marketing Communication Mix. Management Science 2008;54(3):477-491.
38. Kissler J, Herbert C, Peyk P, Junghofer M. Buzzwords: early cortical responses to emotional words during reading. Psychol Sci 2007 Jun;18(6):475-480. [doi: 10.1111/j.1467-9280.2007.01924.x] [Medline: 17576257]
39. Bayer M, Sommer W, Schacht A. Font size matters--emotion and attention in cortical responses to written words. PLoS One 2012 May;7(5):e36042 [FREE Full text] [doi: 10.1371/journal.pone.0036042] [Medline: 22590518]
40. Smith S, Petty R. Message Framing and Persuasion: A Message Processing Analysis. Personality and Social Psychology Bulletin 1996 Mar 01;22(3):257-268. [doi: 10.1177/0146167296223004]
41. Kahneman D, Tversky A. Prospect Theory: An Analysis of Decision under Risk. Econometrica 1979 Mar;47(2):263-292. [doi: 10.2307/1914185]
42. Berger J. Arousal increases social transmission of information. Psychol Sci 2011 Jul;22(7):891-893. [doi: 10.1177/0956797611413294] [Medline: 21690315]
43. Dang-Xuan L, Stieglitz S. Impact and Diffusion of Sentiment in Political Communication-An Empirical Analysis of Political Weblogs. 2012 Presented at: Sixth International AAAI Conference on Weblogs and Social Media; 2012; Dublin, Ireland.
44. Huffaker D. Dimensions of Leadership and Social Influence in Online Communities. Human Communication Research 2010;36(4):593-617. [doi: 10.1111/j.1468-2958.2010.01390.x]
45. Joyce E, Kraut R. Predicting Continued Participation in Newsgroups. J Comp Mediated Comm 2006 Apr;11(3):723-747. [doi: 10.1111/j.1083-6101.2006.00033.x]
46. McKnight H. Social Networking Information Disclosure and Continuance Intention: A Disconnect. 2011 Presented at: 44th Hawaii International Conference on System Sciences; 2011; Manoa, HI p. 1-10. [doi: 10.1109/HICSS.2011.379]
47. Mechanic D. Trust and Informed Consent to Rationing. The Milbank Quarterly 1994;72(2):217. [doi: 10.2307/3350294]
48. Mechanic D. The Functions and Limitations of Trust in the Provision of Medical Care. Journal of Health Politics, Policy and Law 1998 Jan 01;23(4):661-686. [doi: 10.1215/03616878-23-4-661]
49. Dinev T, Hart P. Internet Privacy Concerns and Social Awareness as Determinants of Intention to Transact. International Journal of Electronic Commerce 2006 Jan 1;10(2):7-29. [doi: 10.2753/JEC1086-4415100201]
50. Staggers N, Norcio A. Mental models: concepts for human-computer interaction research. International Journal of Man-Machine Studies 1993 Apr;38(4):587-605. [doi: 10.1006/imms.1993.1028]
51. Kittur A, Peters A, Diriye A, Bove M. Standing on the schemas of giants: socially augmented information foraging. 2014 Presented at: 17th ACM conference on Computer Supported Cooperative Work & Social Computing; 2014; Baltimore, MD p. 999-1010.
52. Feinberg M, Willer R, Stellar J, Keltner D. The virtues of gossip: reputational information sharing as prosocial behavior. Journal of personality and social psychology 2012;102(5):1015. [doi: 10.1037/a0026650]
53. McElroy T, Keith D. Susceptibility to anchoring effects: How openness-to-experience influences responses to anchoring cues. Judgment and Decision Making 2007;2(1):48-53.
54. Shiv B, Edell Britton J, Payne J. Does Elaboration Increase or Decrease the Effectiveness of Negatively versus Positively Framed Messages? J Consum Res 2004 Jun 01;31(1):199-208. [doi: 10.1086/383435]
55. Nov O, Arazy O. Asymmetric Recommendations: The Interacting Effects of Social Ratings? Direction and Strength on Users. 2015 Presented at: 9th ACM Conference on Recommender Systems; 2015; Vienna, Austria p. 249-252.
56. Amrisha V, Grossmann T, Woodward A. Not all emotions are created equal: The negativity bias in social-emotional development. Psychological bulletin 2008:383-403. [doi: 10.1037/0033-2909.134.3.383]
57. Bansal G, Zahedi F, Gefen D. Moderating Influence of Privacy Concern on the Efficacy of Privacy Assurance Mechanisms for Building Trust: A Multiple-Context Investigation. 2008 Presented at: International Conference on Information Systems; 2008; Paris, France.
58. White M, Pahl S, Buehner M, Haye A. Trust in Risky Messages: The Role of Prior Attitudes. Risk Analysis 2003 Aug;23(4):717-726. [doi: 10.1111/1539-6924.00350]
59. Petty R, Cacioppo J. Communication and persuasion: Central and peripheral routes to attitude change. New York, NY: Springer-Verlag; 1986.
60. Paolacci G, Chandler J, Ipeirotis P. Running Experiments on Amazon Mechanical Turk. Judgment and Decision Making 2010;5(5):411-419.
61. Bartneck C, Duenser A, Moltchanova E, Zawieska K. Comparing the similarity of responses received from studies in Amazon's Mechanical Turk to studies conducted online and with direct recruitment. PLoS One 2015 Apr;10(4):e0121595 [FREE Full text] [doi: 10.1371/journal.pone.0121595] [Medline: 25876027]




XSL•FO
RenderX

62. Buchanan T, Paine C, Joinson A, Reips U. Development of measures of online privacy concern and protection for use on the Internet. J Am Soc Inf Sci 2006 Jan 15;58(2):157-165. [doi: 10.1002/asi.20459]
63. Westin A, Maurici D. E-commerce & privacy: What net users want. Hackensack, NJ: Privacy & American Business; 1998.
64. GET-Evidence. About. 2016. URL: http://evidence.pgp-hms.org/about [accessed 2016-07-13] [WebCite Cache ID 6j7T4Owjx]
65. Office for Human Research Protections. Informed Consent Checklist. 2016. URL: http://www.hhs.gov/ohrp/regulations-and-policy/guidance/checklists/index.html [accessed 2016-07-06] [WebCite Cache ID 6inj7ZLMe]
66. Personal Genome Project. Participation Documents. URL: http://www.personalgenomes.org/harvard/sign-up [accessed 2016-07-06] [WebCite Cache ID 6injSZTb3]
67. 23andMe. Genetic testing for ancestry. URL: https://www.23andme.com/about/consent/ [accessed 2016-07-06] [WebCite Cache ID 6injVgRTc]
68. Joffe S, Cook E, Cleary P, Clark J, Weeks J. Quality of Informed Consent: a New Measure of Understanding Among Research Subjects. JNCI Journal of the National Cancer Institute 2001 Jan 17;93(2):139-147. [doi: 10.1093/jnci/93.2.139]
69. Dunn L, Jeste D. Enhancing informed consent for research and treatment. Neuropsychopharmacology 2001 Jun;24(6):595-607 [FREE Full text] [doi: 10.1016/S0893-133X(00)00218-9] [Medline: 11331139]
70. Riegelsberger J, Sasse M, McCarthy J. The researcher's dilemma: evaluating trust in computer-mediated communication. International Journal of Human-Computer Studies 2003 Jun;58(6):759-781. [doi: 10.1016/S1071-5819(03)00042-9]
71. Sepucha K, Fagerlin A, Couper M, Levin C, Singer E, Zikmund-Fisher B. How does feeling informed relate to being informed? The DECISIONS survey. Med Decis Making 2010 Sep;30(5 Suppl):77S-84S. [doi: 10.1177/0272989X10379647] [Medline: 20881156]
72. Moorman P, Newman B, Millikan R, Tse C, Sandler D. Participation Rates in a Case-Control Study. Annals of Epidemiology 1999 Apr;9(3):188-195. [doi: 10.1016/S1047-2797(98)00057-X]
73. Howell D. Statistical Methods for Psychology. Belmont, CA: Wadsworth Publishing; 2013.
74. Farzan R, Dabbish L, Kraut R, Postmes T. Increasing commitment to online communities by designing for social presence. 2011 Presented at: ACM conference on Computer-Supported Cooperative Work; 2011; Hangzhou, China.
75. Baumer E, Katz S, Freeman J, Adams P, Gonzales A, Pollak J, et al. Prescriptive persuasion and open-ended social awareness: expanding the design space of mobile health. 2012 Presented at: ACM Conference on Computer Supported Cooperative Work; 2012; Seattle, WA p. 475-484.
76. Burstein F. In: Holsapple CW, editor. Handbook on Decision Support Systems 1: Basic Themes (International Handbooks on Information Systems). Berlin, Germany: Springer; 2008:121-140.
77. Diakopoulos N, Shamma D. Characterizing debate performance via aggregated twitter sentiment. 2010 Presented at: ACM Conference on Human Factors in Computing Systems; 2010; Atlanta, GA p. 1195-1198.


## Abbreviations

**DTCGT:** direct-to-consumer genetic testing


*Edited by G Eysenbach; submitted 17.02.16; peer-reviewed by H McKnight, M Ptaszynski; comments to author 16.03.16; revised version received 12.04.16; accepted 24.06.16; published 20.07.16.*

*Please cite as:*
*Balestra M, Shaer O, Okerlund J, Westendorf L, Ball M, Nov O*
*Social Annotation Valence: The Impact on Online Informed Consent Beliefs and Behavior*
*J Med Internet Res 2016;18(7):e197*
*URL: http://www.jmir.org/2016/7/e197/*
*doi:10.2196/jmir.5662*
*PMID:27439320*


©Martina Balestra, Orit Shaer, Johanna Okerlund, Lauren Westendorf, Madeleine Ball, Oded Nov. Originally published in the Journal of Medical Internet Research (http://www.jmir.org), 20.07.2016. This is an open-access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/2.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on http://www.jmir.org/, as well as this copyright and license information must be included.



XSL•FO
RenderX

Original Paper

# A Social Media mHealth Solution to Address the Needs of Dengue Prevention and Management in Sri Lanka

May O Lwin[1*], PhD; Santosh Vijaykumar[2*], PhD; Vajira Sampath Rathnayake[1], BSc, CompSci; Gentatsu Lim[1], BA(Hons); Chitra Panchapakesan[1], MSc; Schubert Foo[1], PhD; Ruwan Wijayamuni[3], MPH; Prasad Wimalaratne[4], PhD; Owen Noel Newton Fernando[2], PhD

[1]Centre of Social Media Innovations for Communities, Nanyang Technological University, Singapore, Singapore

[2]Nanyang Technological University, School of Computer Science and Engineering, Singapore, Singapore

[3]Colombo Municipal Council, Colombo, Sri Lanka

[4]University of Colombo, School of Computing, Colombo, Sri Lanka

[*]these authors contributed equally

**Corresponding Author:**
Santosh Vijaykumar, PhD
Nanyang Technological University
School of Computer Science and Engineering
50 Nanyang Avenue, North Spine, N4-02c-75
Singapore, 639798
Singapore
Phone: 65 67905783
Fax: 65 67926559
Email: santoshv@ntu.edu.sg

## Abstract

**Background:** Sri Lanka has witnessed a series of dengue epidemics over the past five years, with the western province, home to the political capital of Colombo, bearing more than half of the dengue burden. Existing dengue monitoring prevention programs are exhausted as public health inspectors (PHIs) cope with increasing workloads and paper-based modes of surveillance and education, characterizing a reactive system unable to cope with the enormity of the problem. On the other hand, the unprecedented proliferation and affordability of mobile phones since 2009 and a supportive political climate have thus far remained unexploited for the use of mobile-based interventions for dengue management.

**Objective:** To conduct a needs assessment of PHIs in Colombo with respect to their dengue-related tasks and develop a new mobile-based system to address these needs while strengthening existing systems.

**Methods:** One-on-one in-depth interviews were conducted with 29 PHIs to a) gain a nuanced, in-depth understanding of the current state of surveillance practices, b) understand the logistical, technological and social challenges they confront, and c) identify opportunities for mobile-based interventions. Quantitative analysis included simple descriptive statistics while qualitative analysis comprised textual analysis of 209 pages of transcripts (or nearly 600 minutes of conversations) using grounded theory approaches.

**Results:** Current paper-based data collection practices for dengue surveillance involved a circuitous, time consuming process that could take between 7-10 days to officially report and record a single case. PHIs confronted challenges in terms of unreliable, standalone GIS devices, delays in registering mosquito breeding sites and lack of engagement from communities while delivering dengue education. These findings, in concert with a high motivation to use mobile-based systems, informed the development of Mo-Buzz, a mobile-based system that integrates three components – digitized surveillance, dynamic disease mapping and digitized dengue education – on a common platform. The system was developed through an iterative, evolutionary, collaborative process, consistent with the Spiral model of software development and is currently being used by all 55 PHIs in the CMC system.

**Conclusions:** Given the entrenched nature of existing paper-based systems in PHIs' work habits, we expect a gradual adoption curve for Mo-Buzz in the future. Equally, we expect variable adoption of the system with respect to its specific components, and specific PHI sub-groups (younger versus older). The Mo-Buzz intervention is a response to multiple calls by the global mHealth community for collaborations in the area of mobile interventions for global health. Our experience revealed that the benefits of this paradigm lies in alleviating country-specific public health challenges through a commonly shared understanding of cultural



XSL•FO
RenderX

mores, and sharing of knowledge and technologies. We call upon future researchers to further dissect the applicability of the Spiral Model of software development to mHealth interventions and contribute to the mHealth evidence debate from theoretical and applied perspectives.

***(J Med Internet Res 2016;18(7):e149)*** doi:10.2196/jmir.4657

**KEYWORDS**

dengue; public health inspector; mhealth; social media; surveillance; needs assessment; prevention and management

## Introduction

Dengue, the vector borne disease that threatens the lives of millions of people in tropical countries, has severely affected Sri Lanka in the past 2 decades. In 2014, the country reported nearly 40,000 dengue cases, a level of burden that has been consistent over the past few years [1]. Curiously, more than 55% of such cases were found to originate from the western province of Colombo, where the country's political capital is located.

Given that the severity of dengue outbreaks has failed to abate, Sri Lanka, and more specifically, the capital city of Colombo, grapples with an exhausted dengue outbreak management system. Dimensions of the systemic fatigue are sporadically highlighted in the mainstream media, such as the case of the public health inspectors (PHIs)—the last mile in Sri Lanka's public health delivery system—who are overtly burdened to an extent of one PHI covering a population of nearly 50,000 citizens [2]. Beyond dengue surveillance, the PHI's daily duties include contributing to the control of other communicable and noncommunicable diseases, reporting on housing and sanitation issues, water supply and waste control, adolescent and reproductive health, and health education and promotion among others. It is reasonable to assume that this wide-ranging job description places undue demands on, and adversely affects, the efficacy of the dengue management system in Colombo. To our knowledge, no research studies have examined occupational challenges faced by PHIs in Colombo, as there have been on rural health workers in India, Vietnam, and sub-Saharan Africa [3-8]. By similar accounts, there is limited evidence critically examining their specific role in the vector management system in a way that gives us a glimpse into opportunities for potential interventions to enhance the effectiveness of prevention programs.

### Role of Mobile Technology in Dengue Prevention

Sri Lanka has witnessed an unprecedented growth in the penetration of mobile services after 2009 when the civil war ended. Currently, Sri Lanka boasts one of the most affordable rates of mobile services across the world, with penetration rates higher than most developing countries [9]. These developments are reflected in national corporate and governmental policies that have together initiated a series of Mobile for Development (or M4D) programs with health and education serving as priority areas [10]. However, dengue programs have yet to benefit from this technological trend, even as vast swathes of the Sri Lankan population become increasingly susceptible to this vector-borne disease.

In other developing countries in tropical regions, technological interventions for bolstering dengue surveillance have mainly focused on the use of geographical information systems (GISs) and other surveillance systems to facilitate early notification or warnings of potential outbreaks. For instance, Chang and colleagues [11] used Google Earth and ArcGIS 9 to create a surveillance system in Nicaragua that can allow public health workers to identify high indices of mosquito infestation in relation to larval development sites like garbage piles and stagnant water pools. In Brazil, researchers developed the SMCP-Aedes, an entomological surveillance system focused on collecting, storing, analyzing, and disseminating mosquito-related information on the Web [12]. In Thailand, Ditsuwan and colleagues [13] used a combination of a national surveillance system database and GIS to evaluate the burden of dengue and chikungunya fever. Dengue-GIS has also been used for monitoring and evaluating national-level epidemiological, entomological, and control interventions in Mexico and has been found to be useful for decision making at different levels of the dengue control system [14]. Although these initiatives have attempted to use GIS for different aspects of dengue prevention and control programs in their respective countries, we recognize 3 main limitations in extant work. First, we notice a paucity of technological interventions that reach beyond the offices of health policymakers and authorities to influence the actual workflow of health workers at points where they interact with the public. Second, it is evident that most technological interventions are focused mainly on surveillance but rarely facilitate efficient health education or community engagement, two of the bedrocks of any powerful dengue prevention program [15,16]. Third, the cost-effective nature and multifunctional capabilities of mobile phones have been used for a range of public health concerns in developing countries [17], but not as much for dengue.

On the basis of the previously mentioned review of literature, our exploratory study was guided by the following aims: (1) to gain a comprehensive understanding of the epidemiological process of identifying, reporting, and recording dengue cases in Colombo, (2) to understand the PHI's logistical, technological, and social risks and challenges in the processes identified in (1), (3) to identify opportunities for technological intervention based on existing beliefs about and familiarity with mobile technologies among PHIs, and (4) to develop a technological intervention that can address the most critical gaps in their existing workflow to enhance the overall efficiency of the dengue management system in Colombo.

Our paper is presented in 3 main sections. First, we present results from a mixed-methods technological needs assessment of PHIs in Colombo. Second, we present a detailed description



XSL•FO
RenderX

of a social media–based system, called Mo-Buzz, which we developed to address the most critical bottlenecks in the current *paper-based* dengue information system. The final section culminates with a discussion of study findings, implications of such a system for the larger public health infrastructure in Colombo, and potential future research.

## Methods

To assess the dengue-related informational and technological needs of the PHIs in Colombo, we conducted a series of in-depth interviews with them that would allow us to gain a nuanced, multifaceted perspective in the issues of utmost concern to them. Each interview was preceded by a simple quantitative survey. We next present details of each of the survey and the in-depth interviews.

### Survey

The survey was designed to obtain a demographic profile of the PHIs; examine their technological habits, usage, and preferences in the dengue context; and generally assess their readiness to move forward and adopt and use the technology. The questionnaire comprised sections on measurement of demographic variables such as age, ethnicity, education, income, and years of experience as a PHI. We then captured technology use through a simple yes or no question asking whether they have previously used the Internet, simple mobile phone, a smartphone, mobile apps, and a tablet. Descriptive analyses were conducted using univariate statistical techniques on SPSS, v. 21 (IBM Corp. Armonk, NY).

### In-Depth Interviews

The in-depth interviews complemented the survey and were aimed at getting a deeper, more detailed, and nuanced perspective on the PHIs role in the public health system, their specific functions related to dengue prevention and management, and their beliefs on the potential for technological interventions in their dengue-related duties. An interview guide facilitated the flow of the conversation across the following themes: (1) roles and responsibilities of the PHI, (2) perspectives on the dengue burden in Colombo, (3) data collection and information flow pertaining to dengue monitoring and surveillance and challenges in this process, (4) health educational activities related to dengue, (5) technology use and preferences, (6) ideas for technological intervention, and (7) perspectives on client (the general public or community members whom the PHIs interface with on a daily basis) interaction, client trust, and client satisfaction. All interviews were conducted in Sinhalese or Tamil (the major local languages). All interviews were digitally recorded and later translated into English by an experienced translator.

Data were collected through a series of detailed one-on-one interviews with the PHIs at the office of the Colombo Municipal Council (CMC). All interviews were conducted with the permission and approval of the Chief Medical Officer of the CMC who also assured access to the PHIs. The Chief Medical Officer briefed the cadre of PHIs about the interviews (before the commencement of the study) and set up a schedule where every PHI would be scheduled to attend a 30- to 45-minute time slot at the interview venue based on their schedule. At the start of each session, the research staff described the aims and purpose of the study to the PHI and obtained their signatures on an informed consent form after explaining the terms of the study including data confidentiality. Each study comprised a short 10-minute quantitative component followed by an in-depth interview (qualitative) that lasted between 30 and 60 minutes.

Qualitative data analysis was conducted through grounded theory approach in 2 ways. First, analysis proceeded as data were collected, thereby allowing early findings to influence later inquiries. For instance, when we observed that the first few participants alluded to contextual constraints associated with dengue-related data collection on the ground (a subtopic not originally part of our interview guide), we gradually introduced this theme into subsequent interviews to explore this area even more. Second, emergent findings and a review of the transcripts were used to design a preliminary coding guide that was further refined as coding and analysis proceeded. Two researchers coded each transcript independently, and then arrived at a final code after discussing their codes with an adjudicator. The final codes were then processed through NVivo, a software that helped to summarize qualitative data using visual and tabular formats, analyzing frequency and prominence of topics discussed by the PHIs. Eventually, the research team coded 585 minutes of interviews spread over 5 days and coded 209 pages of interview transcripts.

## Results

### Survey

As seen in Table 1, the 29 PHIs whom we interviewed were nearly equally distributed between the 21-30 years and 31-40 years age groups with all but one belonging to Sinhalese ethnicity. Nearly 86% (25/29) of them had attained a diploma. Nearly 45% (13/29) of the PHIs had served their role for less than a year, nearly 21% (6/29) between 1 and 5 years and the remaining 35% (10/29) were regarded as seniors having served more than 5 years.



XSL•FO
RenderX

**Table 1.** Demographic breakdown of the PHIs.

| Category | | Frequency (N) | Percentage (%) |
|---|---|---|---|
| Gender | | | |
| | Male | 29 | 100.0 |
| Age | | | |
| | 21-30 | 16 | 55 |
| | 31-40 | 13 | 45 |
| Ethnicity | | | |
| | Sinhalese | 28 | 96 |
| | Indian Tamil | 1 | 4 |
| Highest educational level | | | |
| | Secondary and below | 3 | 10 |
| | Certificate or diploma | 25 | 86 |
| | University and above | 1 | 4 |
| Duration of service | | | |
| | Less than 1 year | 13 | 45 |
| | 1-5 years | 6 | 21 |
| | More than 5 years | 10 | 35 |
| History of digital technology use | | | |
| | Internet | 25 | 86 |
| | Simple mobile phones | 26 | 90 |
| | Smartphones | 18 | 62 |
| | Mobile apps | 10 | 35 |
| | Tablets | 8 | 28 |

Our descriptive analysis revealed a healthy history of technology use with nearly 86% (24 of 28) of the PHIs having used the Internet and 90% (25 of 28) having used simple mobile phones. Of these, only 62% (17 of 28) of them had used smartphones previously, and a mere 28% (8 of 28) had prior experience with tablets. Finally, nearly 35% (10 of 28) of the PHIs were familiar with mobile apps.

An analysis of the constructs (Table 2) revealed interesting insights. Although perceived ease of using [18] mobile apps and tablets was relatively low with means of 3.63 and 3.47, respectively, their perceived usefulness [18] for dengue collection was among the highest, with means of 3.52 and 3.86, respectively. In addition, the PHIs reported that better technology would strengthen their ability to track (mean [M]=4.52, standard deviation [SD]=.63) and report (M=4.69, SD=.71) dengue cases more efficiently and make it easier for them to identify new mosquito breeding sites (M=4.28, SD=.92).



XSL•FO
RenderX

**Table 2.** Technology-related attitudes.

| Constructs | | M | SD |
|---|---|---|---|
| Perceived ease of use | | | |
| | Internet | 4.28 | .88 |
| | Simple mobile phones | 4.77 | .62 |
| | Smartphones | 4.34 | .80 |
| | Mobile apps | 3.63 | 1.10 |
| | Tablets | 3.47 | 1.10 |
| Perceived usefulness | | | |
| | Paper and pen | 3.41 | 1.46 |
| | Simple mobile phones | 2.79 | 1.43 |
| | Smartphones | 3.54 | 1.03 |
| | Mobile apps | 3.52 | 1.15 |
| | Tablets | 3.86 | 1.23 |
| Perceived utility for dengue tasks | | | |
| | I can easily track new dengue cases in Colombo | 3.41 | 1.02 |
| | I can easily report new dengue cases in Colombo | 3.61 | 1.01 |
| | I can easily identify new mosquito breeding sites in Colombo | 3.24 | 1.02 |

## In-Depth Interviews

In this section, we first present a description of the existing paper-based surveillance process as reported by the PHIs and then examine specific issues of interest to our technological development process.

### *Understanding How Dengue Cases Are Identified, Reported, and Stored in Colombo*

We obtained a comprehensive understanding of the flow of dengue-related information between different agencies involved in the dengue surveillance programs in Colombo. As shown in Figure 1, the existing dengue information architecture reflects a circuitous and time-consuming process. This process commences with a patient who experiences symptoms visiting the hospital who in turn hand over a paper-based record of suspected dengue cases to the PHI who is assigned to that particular hospital. All PHIs who receive this information pass it along to the CMC Epidemiological Unit (CMC-EU), where an official is assigned to create a separate file for individual patients. The official categorizes all these files according to the Medical Officer of Health (MOH) jurisdiction under which they are covered and dispatches this information to each of the MOH offices through the CMC-EU. The MOHs then distribute the files to their PHIs for follow-up through patient visits. Each PHI visits the patient to confirm his or her diagnosis for dengue, on which a decision is taken to fill either a Communicable Disease Form (CDF) and a Dengue Investigation Form (DIF) or only the former, depending on whether the patient is tested positive or negative. In addition, in case of a positive diagnosis, the PHI is required to conduct a house and area inspection to identify possible mosquito breeding sites and educate the patient and his family on protecting themselves from dengue. After obtaining the entire set of CDFs and DIFs from the PHIs under their jurisdiction, the MOHs officially approve the forms before dispatching them to the CMC-EU. The CMC-EU manually collates the information from all the DIFs to create a record, map dengue cases on a manual map, and ensure that all the cases are within the CMC jurisdiction. At the end of this process, a formal report is sent to the CMC Public Health Department who officially sign on it before dispatching it to the Ministry of Health. The whole process could take anywhere from 7 to 10 days.

We now examine specific issues of interest. Table 3 tabulates the distributions of each topic mentioned by PHIs and their prevalence in the overall discussion.



XSL•FO
RenderX

**Table 3.** Distribution of topics.

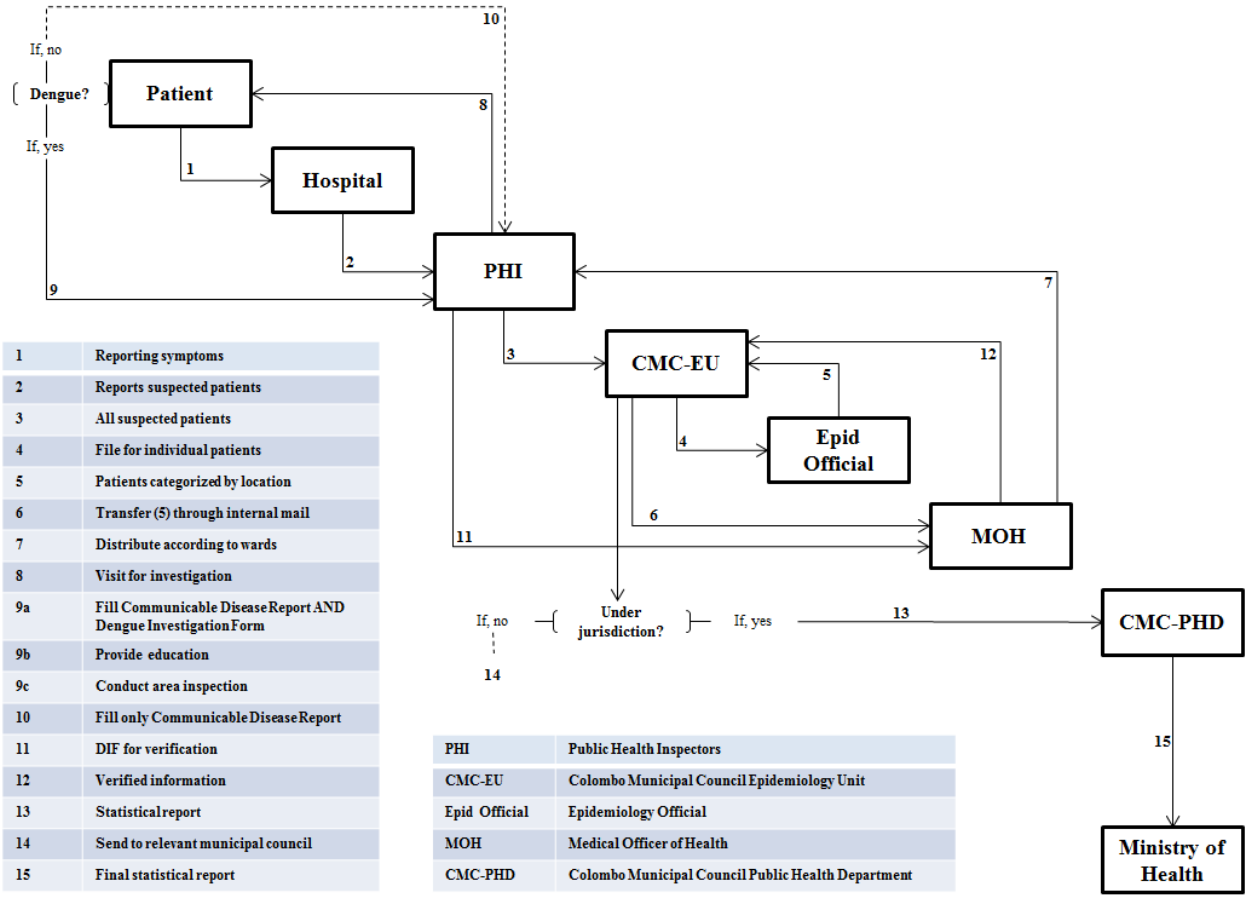
| Topic | Percentage of PHIs who discussed a specific topic (%) | Percentage of reference in relation to overall references (%) | Percentage of reference in relation to overall conversation size (%) |
|---|---|---|---|
| Barriers impeding PHI's work | 100.0 | 13.6 | 16.7 |
| Epidemiology about dengue | 100.0 | 9.4 | 13.6 |
| Process of PHI's work | 100.0 | 19.4 | 21.1 |
| Prevention of dengue | 96.4 | 5.4 | 8.9 |
| Knowledge about dengue (PHIs) | 92.9 | 5.1 | 7.0 |
| Education materials about dengue | 89.3 | 8.0 | 9.3 |
| Attitude of public towards PHIs | 86.0 | 6.4 | 6.0 |
| Equipment used for dengue tasks | 85.7 | 5.6 | 8.0 |
| Profile of PHI | 85.7 | 2.5 | 7.7 |
| Suggestion for mobile app | 85.7 | 6.3 | 7.6 |
| Responsibility of PHIs | 78.6 | 4.4 | 4.5 |
| Facilitators to PHI's work | 71.4 | 2.6 | 3.8 |
| Client interaction with PHIs | 68.0 | 3.0 | 6.0 |
| Burden of dengue | 67.9 | 2.0 | 3.0 |
| Knowledge about dengue (public) | 67.9 | 1.9 | 2.9 |
| Trust among clients on PHIs | 50.0 | 1.4 | 1.4 |
| Story from PHI's work life | 46.4 | 1.0 | 3.5 |
| Diseases related to mosquitoes | 39.3 | 0.9 | 2.8 |
| Technology use of PHIs | 25.0 | 0.4 | 0.5 |
| Demographic factors | 14.3 | 0.4 | 0.4 |
| Client satisfaction | 10.7 | 0.3 | 1.0 |

The interviews revealed that the most complex and challenging sub-processes from the process (presented in steps 8 and 9 of Figure 1) involved PHIs field visits to the residences of potential dengue patients and the follow-up actions. The other steps mostly involved manual transfer of documents from one set of actors to the other within the system, but steps 8 and 9 involved multiple logistical, technological, and social elements that infused tension in the system. As such, these steps are the bedrock of the dengue data surveillance in Colombo and thus required most attention. Here, we outline practical challenges faced by PHIs while executing their dengue data collection tasks.



XSL•FO
RenderX

**Figure 1.** Existing flow of dengue information in Colombo, Sri Lanka.




| | |
|---|---|
| 1 | Reporting symptoms |
| 2 | Reports suspected patients |
| 3 | All suspected patients |
| 4 | File for individual patients |
| 5 | Patients categorized by location |
| 6 | Transfer (5) through internal mail |
| 7 | Distribute according to wards |
| 8 | Visit for investigation |
| 9a | Fill Communicable Disease Report AND Dengue Investigation Form |
| 9b | Provide education |
| 9c | Conduct area inspection |
| 10 | Fill only Communicable Disease Report |
| 11 | DIF for verification |
| 12 | Verified information |
| 13 | Statistical report |
| 14 | Send to relevant municipal council |
| 15 | Final statistical report |

| | |
|---|---|
| PHI | Public Health Inspectors |
| CMC-EU | Colombo Municipal Council Epidemiology Unit |
| Epid Official | Epidemiology Official |
| MOH | Medical Officer of Health |
| CMC-PHD | Colombo Municipal Council Public Health Department |

### *Logistical Challenges*

The foremost uncertainty for the PHIs arose when they would attempt to visit the residences of the prospective dengue patients, as assigned by their respective MOHs. The PHIs noted multiple instances when the patients would either be away or would have relocated, rendering their visit futile. In some cases where it would be challenging for the PHI to locate the client's exact address, the PHI would be unable to contact them and inquire as the patient would have refused to share their mobile number on account of privacy issues. If the clients were available, the PHIs would have to complete the requisite procedures, fill up the lengthy forms, and then commute to the respective office to submit them. A combination of these factors contributed to delays in data collection ultimately reducing systemic efficiency. The laboriousness of the process was highlighted by PHI 23 when he noted:

> *I would say finding the patients is the biggest challenge. Firstly, we don't have the required transport. Sometimes a person's address has been falsely stated. So even though they are there in that area we cannot find them. Some are reluctant to share this information with us. I'm not sure why but even when we call and ask them they do not say what their true address is.*

At times, the time taken to complete the procedures preceding the PHIs visits would adversely affect the surveillance process as the risk factors would have become obsolete or the patients have recovered.

In the words of PHI 12:

> *…50% there is a delay, mostly in the sending of reports. When we get the report the breeding places may not exist anymore (and) maybe the patient might have been relocated after getting better even (sic).*

An integral procedure followed by the PHIs during their client visits is to engage them and obtain details for completing 2 forms: the CDF and the DIF, which are both World Health Organization–approved protocols. The sheer length of these forms, which commanded anywhere between 20 and 45 minutes of the PHI's visits time, apart from its paper-based format requiring manual entry made these procedures prone to error and logistical inconvenience.

Commenting on the imminent risks of losing critical health data from these forms, PHI 13 noted:

> *We need a lot of space to store these…and if some physical damage happens like mice getting to the forms, the information can be lost.*

Finally, PHIs were educating dengue-affected individuals and families using outmoded means of health communication such as pamphlets and brochures. With minimum persuasive impact and lack of audience engagement, PHI reported that these



XSL•FO
RenderX

materials bore minimum effects of attitudes, knowledge, and behaviors related to dengue.

### *Technological Challenges*

An integral part of the PHIs' duties during client visits is identification of breeding sites and recording information about them that can be used for follow-up actions such as fogging. However, the difficulty of the dense and unstructured urban terrain in and around Colombo meant that the geographical information recorded by the PHIs suffered on accounts of accessibility and accuracy.

In 2011, the CMC initiated a technological intervention to address this problem by collaborating with the WHO to procure Geographical Positioning System (GPS) devices. The PHIs would thus be required to carry these devices with them in addition to the rest of the paperwork, and after detecting the particular locality's geographical coordinates, note it down on their paper forms. Although these devices provided partial reprieve to the PHIs, this intervention started triggering its own set of unique technical challenges. For instance, inconsistencies in recording the geographical locations were bound to occur, as each MOH (overseeing multiple PHIs) was equipped with only one device, resulting in a situation where not all PHIs could be uniformly equipped with the system. Furthermore, the density of the terrain comprising uneven housing patterns, unplanned road layouts, and slum colonies meant that the GPS device was, in many instances, unable to procure the signal required to display the precise coordinate of breeding sites. PHI 9 explained the situation thus:

> *Sometimes I can't get GPS points. … Sometimes, we have to go like 50 meters away in order to get a significant difference in the GPS coordinates...*

In addition, the back-end data management system for the GPS device was set up in a way that required the coordinates captured to transmit through multiple points before finally getting recorded, thereby causing inordinate delays in the data collection process. Pointing out the disadvantage of such delays to breeding site surveillance efforts, PHI 14 noted:

> *When we give GPS points they form clusters on the central map. Then we can pinpoint breeding places on the map. When we do this in the current system the data has to go here and there and the delay may take days and the breeding place will have served its purpose already.*

Commenting on the inability of the existing GPS systems to equip the PHIs with visual maps that can inform them about where dengue outbreaks are occurring, PHI 21 said:

> *…we don't know how mapping must be done exactly. If it's with us then we can know which areas are more prone to cases. But it's just that we refer record books. We don't get the mapping information. Although we get the waypoints, we don't get the resulting mapped data. That does not come to us. We have no feedback from this.*

Public health regulations in Sri Lanka allow for legal action to be taken against offenders who fail to address the problem of breeding sites in and around their homes or construction sites (in case of faulty builders). Although the current arrangements allowed the PHIs to record and report the geographical coordinates of errant offenders, the PHIs were unable to provide photographic evidence that would bolster the implementation of such punitive actions. Although some PHIs on their initiatives used their phone camera to capture pictures of breeding sites, these images were seldom allowed as official evidence in the court. Compounding the problem was the fact that not all PHIs were equipped with camera phones, thereby potentiating a situation of inconsistent evidence from the health authorities to the courts.

### *Social Challenges*

Despite serving as the last and most critical mile of the public health care system in Sri Lanka, PHIs invoke a range of reactions from the communities they serve during client visits. Ranging from fear and caution to resistance and apathy, these reactions can sometimes stymie the efficiency and speed with which they implement their tasks, as it influences the quality of interaction between the PHI and the client in different ways. At the entry stage, when the PHIs have arrived at the clients' residence and are introducing themselves, they are used to being met with a range of responses. Some clients immediately request the PHIs to take a seat and are keen to share all the information that the latter need so that their disease can be cured, whereas others could make the PHIs feel unwelcome. As PHI 14 says, over time PHIs have developed their own strategies to smoothen their interaction with the client.

> *After we go to a patient we can't suddenly start to fill the form. We must become friendly with them. Then only can we get the correct answers from them.*

The PHIs' affable approach can yield limited results especially when clients are requested for their income as part of the demographic assessment in the forms. The PHIs concur that while this question is casually dealt with by residents in lower income neighborhoods; clients in higher income neighborhoods are more cautious and are thus reluctant to share such information. According to PHIs, client attitudes come into the fore especially when it comes to communicating to them about the actual risks of dengue.

PHI 24 shared his experience thus:

> *Sometimes, when I tell the truth it's like a joke to them. When we come they say "here comes the dengue (sic)." They take the leaflet and just throw it somewhere. They don't want to read or they don't like it. They think it's a joke, what we're doing.*

A number of the PHIs reported that although they frequently complemented the information in the leaflets with verbal explanations about modes of transmission about dengue and how clients could protect themselves. They were confounded by denial and apathy, especially by members from a particular ethnic community who would, at times, not even allow the PHIs into their houses.

In the words of PHI 9:



XSL•FO
RenderX

> *…I told them 'You know more than us about how to stop the spread of dengue through the media and all. Yet you cannot feel the shock you feel now when one of your children is at the risk of death.' I asked them 'Do you want to hear that your child is dying?' Then they got frightened.*

As part of their community education efforts, the CMC had also created a movie with an attempt to use the power of an audiovisual medium to enhance the appeal of the messaging. However, the PHIs reported that it was challenging to gather community members at one site at which the movie could be played, and that the contents of the movie too were outdated. In essence, although PHIs acknowledged the potential of moving images to better communicate risk messages about dengue, they equally recognized the inherent constraints in effectively bringing this strategy to the public. PHI 21 elaborated on these challenges:

> *When we show the movie it changes from area to area. If it's Cinnamon Gardens who can I show the movie to? …. They aren't interesting and have no storyline. Also, the people in shanties have no time to spend watching this. They are always trying to spend their time to find some money. So watching a film on dengue is the least of their priorities.*

### *Identify Opportunities for Technological Intervention Based on Existing Beliefs and Attitudes and Technological Habits and Exposure*

As seen previously, our brief survey preceding the interviews had revealed that despite the limited exposure to and experience with mobile phones, the PHIs demonstrated positive attitudes toward these technologies if integrated into their work. The in-depth interviews helped to generate a more nuanced understanding of the rationale behind their attitudes and the multiple ways in which these technologies could ease their work life, while bolstering dengue prevention efforts in Colombo.

On the basis of their knowledge and our description of mobile phone and tablet capabilities, the PHIs concurred that these technologies would assist in facilitating the data transmission, reporting, and collecting processes. Specifically, the PHIs believed that the burden of executing their dengue-related tasks would be substantively alleviated as these technologies could help PHIs to obtain accurate addresses, dispatch reports from the field rapidly, and possibly even obtain maps of dengue cases. These would not only add to the overall efficiency of their daily performance but also fortify their understanding of dengue spread on a real-time basis, useful knowledge that can be transferred to other dengue-related tasks such as health education. Explaining how the Internet-enabled mobile technologies could aid them in responding to and strategizing programs for dengue outbreaks, PHI 12 said:

> *If we can submit our report to the EPID (epidemiology) unit directly from the site, including the breeding site information and location I think it would be very useful to us since the report would have come through immediately. Then we can quickly identify areas which could prove to be severe in the following months.*

In addition to quicker transmission of field reports, other capabilities of mobile phone technologies, such as reminders, could enable PHIs to be reminded of a new patient report that they would be required to follow-up on, thereby allowing them to plan their workdays in advance and in a more time-efficient manner. Many PHIs suggested that one of the main advantages of such technologies would be in delivering dengue education sessions to clients and other community members. PHIs felt that the use of such technologies would be regarded as a social novelty, which, in turn, would arouse curiosity among community members eventually leading to greater engagement. Apart from obtaining a more detailed awareness of dengue, PHIs believed that these technologies bore greater appeal from a social persuasion perspective and could thus help to transform positive attitudes to behavioral performance.

PHI 12 said:

> *Yes, it's more effective definitely. People like it better if we can show them instead of just reciting orally. Even in a school if we demonstrate to the children via a drama for example it becomes more effective.*

In summary, our assessment revealed that the current systemic practices surrounding dengue surveillance and prevention were beset by a number of challenges, some of which could be partially addressed with the aid of mobile technologies. Foremost, we found that the workflow surrounding PHIs' client visits was weighed down by unreliable procurement of clients' addresses, lengthy paper-based form filling prone to environmental risks, and variable accuracy of geographical coordinates recorded by existing GPS devices. After collecting data, the process of transmitting and processing it through various reports was time consuming and adversely affecting the timeliness of surveillance operations. **Finally**, PHIs' delivery of dengue education was stymied by outdated modes of communication such as pamphlets and dated movies. However, the PHIs seemed open to and enthusiastic about an **I**nternet-enabled mobile technological intervention albeit some cautious signals from the older PHIs. Overall, the sentiment we gleaned from the interviews seemed to suggest a number of advantages to such an initiative and the PHIs offered concrete ideas that we could built into our innovation, the details of which are presented in the following section.

## Mo-Buzz: A Socially-Mediated System for Dengue Surveillance, Engagement, and Education

The needs assessment helped to identify the key gaps and constraints in the existing dengue information flow and also opportunities to address these using mobile social media. The challenge was to facilitate easier and more efficient exchange of information between actors without changing the existing workflow that has been established according to national guidelines. Instead of digitally transforming the information flow in its entirety, the priority for our innovation would be to address the bottlenecks in steps 8 and 9 identified in Figure 1 and to facilitate a more effective and efficient client visit by the PHIs. The following sections describe our proposed solution, namely Mo-Buzz, which is a socially mediated system that is



XSL•FO
RenderX

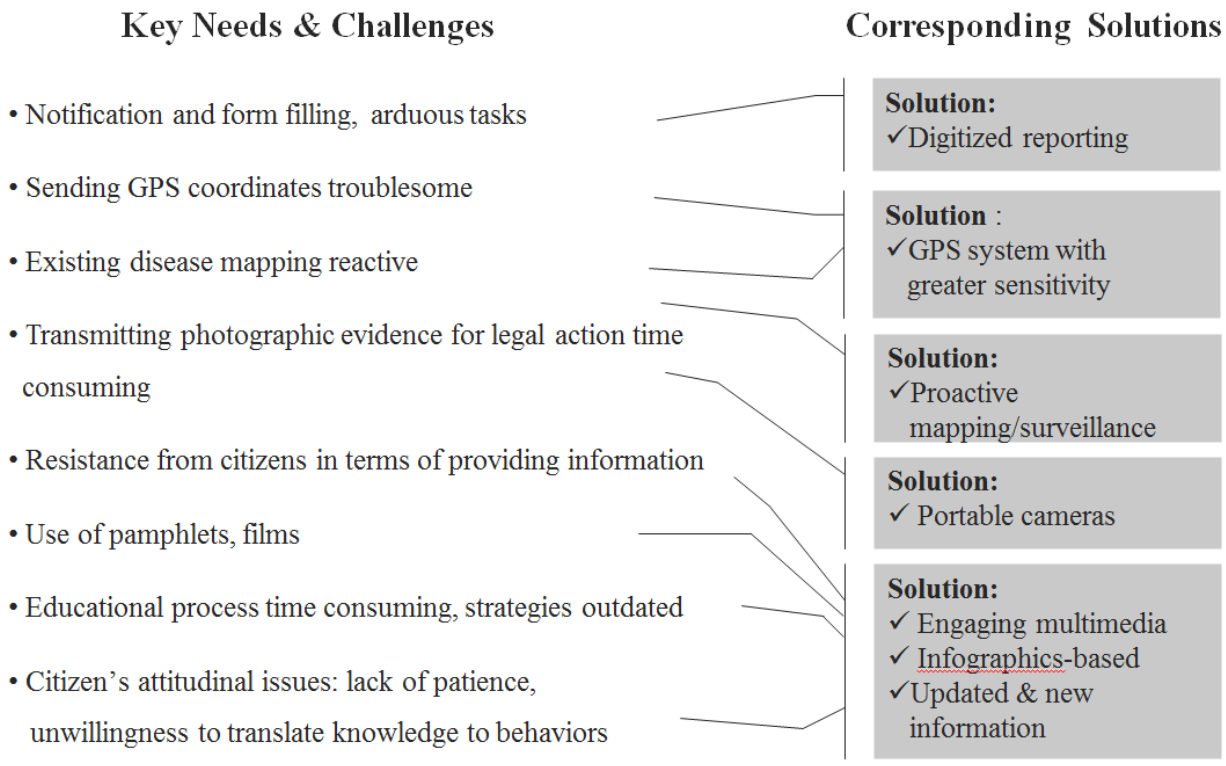
built on integrated information architecture and is available for PHIs on portable tablets.

### Development Approach

Our approach to developing a social media–based solution for addressing the dengue prevention gaps in Colombo was inspired by the Spiral Model of software development [19-21]. This model is premised on a cyclical notion of software development where "risks" attributed to the system might be incrementally reduced through an iterative, evolutionary process of technical refinement that involves concurrent collaboration between multiple stakeholders. Risks are defined as "situations or possible events that can cause a project to fail or meet its goals" [21].

The conceptual alignment of the Spiral Model with our goals for this study are clear as will be demonstrated by an explanation of the model's founding ideas. First, the model recommends that the software development process commence by *determining objectives and identifying constraints*, as has been accomplished by our needs assessment. Second, *risks* in the context of a technology-based solution dengue surveillance and prevention in Colombo—a context where PHIs have minimal exposure to smartphones—might pertain to slow or gradual adoption of the technology by PHIs, the delay involved in managing both, the paper-based forms *and* tablets for a period, and the accuracy of the new technology's reading of the geographical coordinates of a location. Third, involvement of *multiple stakeholders* in the process of developing this technological solution was imminent as any system developed by the research team based in Singapore would need to incorporate the local technical nuances of Colombo, insights that we could best gain by collaborating with institutions such as the CMC, Mobitel (the second largest telecom operator in Sri Lanka), and the University of Colombo School of Computing. Finally, the *evolutionary nature* of technological development was expected because of 3 factors: (1) the system would undergo multiple iterations as PHIs' comfort and familiarity with the technology increased with use over time, (2) the PHIs would receive feedback from their clients on an ongoing basis, which would need to be incorporated into system refinements, and (3) given that the new system would be developed in parallel to the existing system, we would first need to replicate the existing workflow and then attempt to abbreviate and enhance the process in subsequent versions. Figure 2 graphically depicts how the core needs identified through our research were mapped to potential mobile media solutions.

**Figure 2.** Translation of research findings from needs assessment into mobile solutions.




### Technical Specifications

Our system is built on open source technologies and is mainly purposed for mobile and Web-based application which can be accessed through an Android platform (which eventually be extended to include iOS) or a Web browser. The Android solution forms part of the main application by running as an agent on mobile devices. The PHIs and MOH can report information in various forms (photo or text) using mobile devices. The Web-based solution is designed mainly for the management as it offers an interactive system for geospatial visualization, reports for reported DIF, summaries and graphs, and Web forms for other details. The solution is developed using Java-related technologies. The server side of this system is supported by Apache, Tomcat, and MySQL.



XSL•FO
RenderX

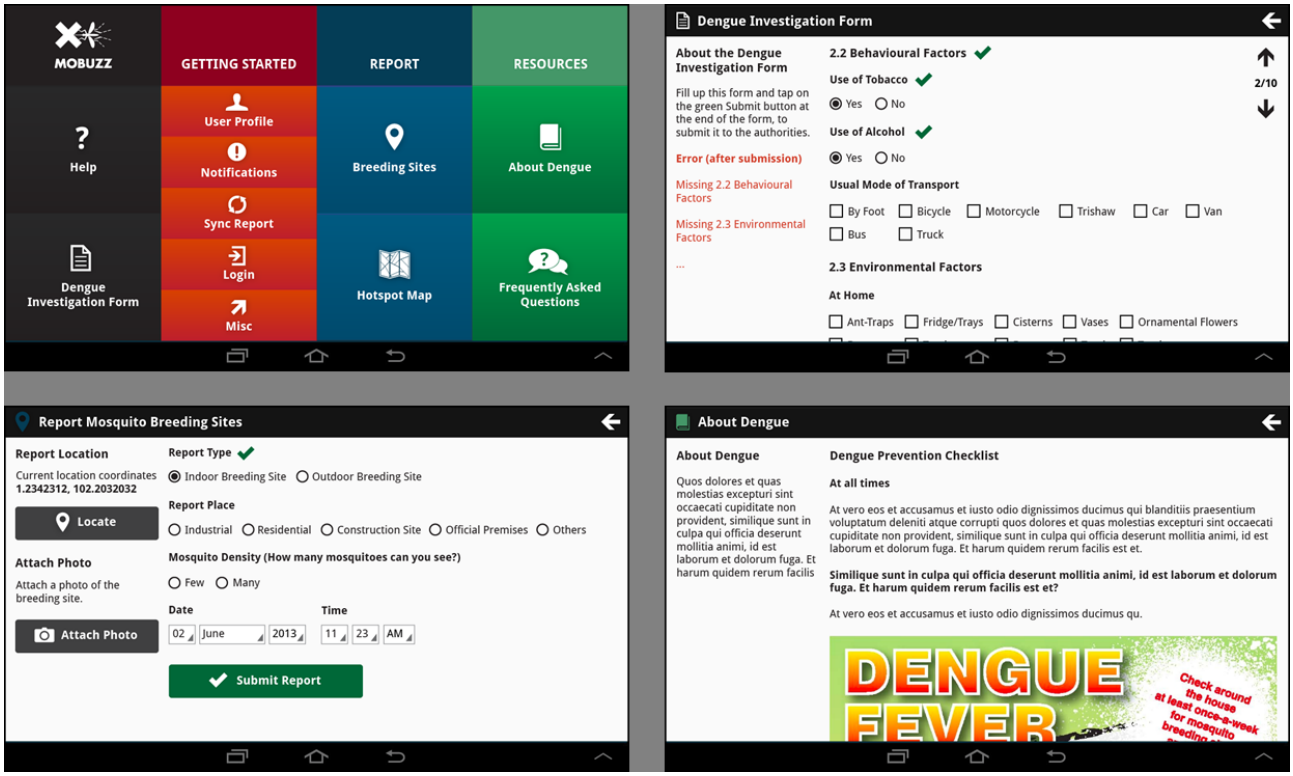
## System Description

The Mo-Buzz system digitizes 3 main functions of PHIs and presents the capability on handheld mobile devices and Web interfaces: (1) capturing, storing and recording visual, textual, and geographical information from patient visits and house or area inspections, (2) staying updated of dengue spread patterns in the Colombo region on a real-time basis, and (3) providing dengue education to the public in an engaging format that will retain their attention and interest.

### *Digital Surveillance*

As seen in Figure 3, this component allows the PHI to capture clients' information on a digitized DIF form, which is easy to use and includes alerts in case the PHI has missed filling out certain fields. The system thus ensures that the DIF forms are not only complete but also are stored for later reference and can be sent to all the relevant authorities in the different agencies (see Figure 1) with the click of a button, thereby drastically reducing reporting time. In addition, the DIFs are automatically linked to Google Map, thus bolstering every individual DIF with accurate geographical coordinates that can be reviewed by the authorities. The main advantage of this functionality is that the authorities can view, on a continuous, real-time basis, the geographical areas from where dengue cases are being reported and take swift action instead of waiting for paper-based reports to arrive in a delayed manner. This component also allows the PHI to capture photographs of breeding sites, which are automatically geotagged, and share it with all relevant authorities in the chain of command to view and take necessary action (such as fogging and pest control).

**Figure 3.** Screenshots from Mo-Buzz depicting the home screen (top left), mosquito reporting form (top right), potential breeding site submission form (bottom left), and health educational component (bottom right).




### *Digitized Dengue Monitoring and Mapping*

In contrast to the CMC's existing manual pin-maps that can only be updated at the end of every case reporting cycle, the Mo-Buzz system offers a live real-time dengue map that is updated as and when PHIs submit a DIF form to the system. This allows the CMC's public managers to obtain real-time updates of dengue spread and allocate dengue prevention and management resources strategically and efficiently. This component also automatically draws information from the geotagged breeding site reports and represents this information visually in a map format so that the MOH and their respective PHIs can plan their prevention activities accordingly.

### *Digitized Dengue Education*

To increase engagement between the PHI and their clients, the Mo-Buzz system offers a tablet-based health education component. This has been done in the backdrop of mounting evidence that suggests positive outcomes resulting from mobile-mediated health education modules for health workers in other contexts. The first version of the health educational module includes digitized versions of the CMC's dengue education materials that the PHI presents to his clients complemented by verbal explanations of dengue prevention concepts. We have enhanced the contents with more graphical elements presented in 3 languages (English, Sinhalese, and Tamil) to create awareness among communities with varying levels of linguistic proficiency. For future versions, we are taking the information we gathered to build enhanced graphics, animations, and tailoring capabilities into the health education component.



XSL•FO
RenderX

### Addressing Risks Through an Iterative Process

Mo-Buzz was developed through a series of iterative steps carried out in continuous collaboration with the CMC management, PHIs, and Mobitel. Consistent with the spirit of the Spiral Model, we chronicle the 3 main "risks" that we encountered and explain how these risks were gradually alleviated in subsequent iterations of the system.

The major challenge for the Mo-Buzz system design was the adoption by its target user-base. Most staff members in CMC have not had an experience with digital devices. At the same time, the daily information collection need was substantial.

Our final system was developed through several versions and tested with stakeholders. Each version's initial scope was selected based on feedback from younger PHIs. The rationale was engaging younger staff members as early adopters and eventually ambassadors before extending to older staff.

User interfaces were designed based on CMC's standard documents and followed similar format for simpler forms to help PHI's to get familiar. They were designed to minimize the navigation depth, reduce tedious typing, and all contexts were grouped according to usage patterns of PHIs. The system was introduced with task-oriented trainings, selecting small groups of staff members based on their roles.

#### *Risks of Technological Adoption and Change Management*

Despite the enthusiasm for mobile technology–based solutions expressed by PHIs and their self-confidence in the ability to handle such solutions, we anticipated challenges in terms of adoption given that our intervention would be completely new to the PHIs' context. As expected, we gradually discovered that while some PHIs displayed lesser technological skills than we expected, the senior PHIs (by age) were resistant to adopt this new system after years of using paper-based methods. We addressed this risk through a 3-pronged, ongoing strategy. First, the research team constantly consulted with PHIs in the process of development, thereby enhancing their familiarity with the system's capabilities, and softening their resistance to adopt it. Second, the research team, in collaboration with the CMC management, conducted a number of training sessions that ingrained in PHIs the advantages of the system, and the mechanisms by which our solution could address their daily dengue surveillance concerns. Third, the previously mentioned 2 strategies allowed us to identify specific resistance points within the technology—for instance, the number of fields in the DIF—that we collaboratively managed to reduce over time, maintaining only the most critical informational fields.

#### *Structural Risks of Technological Implementation*

The technical configuration of our initial system—such as information transfer rate and connectivity strength—was based on our understanding of the strength of mobile technological infrastructure in Colombo, as gleaned from our conversations with experts from CMC and Mobitel, and industry reports. When this version was tested in the controlled environment of Mobitel's offices, we found minimal inconsistencies with our technical expectations. However, when we used the technology in the field (where PHIs would eventually use it), we discovered a number of issues that needed to be ironed out including weak transfer rates and intermittent mobile connectivity. To overcome these problems in a context where PHIs could least afford to lose data from the field, we developed a simple mechanism that would enable them to save completed DIF forms in case they were unable to immediately send it to the CMC due to network connectivity issues. The system would give them options to retry sending or save temporarily and automatically synchronize at a later time when network connectivity was more stable. Similarly, we discovered that it became challenging to send reports of breeding sites with pictures (in an unstable network connectivity environment) as the images comprised large file sizes. The subsequent iterations of Mo-Buzz involved a mechanism that would automatically resize the images to fit the device's screen dimensions without severely compromising on picture quality while reducing file size.

#### *Health Education Materials on Mo-Buzz*

In its incipient stages of development, the Mo-Buzz system comprised digitized versions of existing paper-based health education materials, as desired by the CMC and PHIs. Over time, it became clear that this strategy needed to be revised on account of the simple reason that the multimedia capabilities of tablets were left unused. In subsequent iterations, we collaborated with a team of designers to develop infographics-based health education materials, now available in English, Sinhala, and Tamil and are in the process of incorporating an animation-based dengue education video that the PHIs can show to their community members. We are also in the process of further developing the educational module in a manner that will offer PHIs informational cues and/or alerts in case they fail to cover any important subtopics related to dengue during their interaction with client communities.

We chronicled a range of other risks including issues of linguistic proficiency, range of supported devices, incorporating informational requirements of the CMC management, but have restricted our discussion to 3 risks in the interest of parsimony.

## Discussion

### Principal Findings

Recognizing the severity of the dengue situation in Sri Lanka, our study aimed to identify specific gaps and challenges in surveillance and prevention efforts and use this understanding as the foundation to build an mHealth intervention for PHIs in Colombo. Although our effort was inspired by emerging global efforts using the power of mobile technologies to address public health concerns in developing countries, our objective was to integrate the affordances of multiple solutions and offer them on a common platform, in the form of Mo-Buzz, thereby generating a holistic solution.

Previous studies [22,23] investigating the needs of dengue-related vector-control programs focused on larger systemic issues ranging from insufficient budgets and personnel to challenges in community engagement and interagency collaboration. These findings influenced our intervention to the extent that it fortified our understanding about the potential



XSL•FO
RenderX

constraints and organizational challenges that we were likely to confront. The strength of our intervention, however, lies in the ability to acknowledge that while it would be unrealistic to expect mobile technologies to engender a complete systemic transformation, its greater value would lie in being introduced at specific points in the information flow where the tension was greatest. As a result, Mo-Buzz has been designed specifically to address the bottlenecks in steps 8 and 9, where the human elements of communication and action create unique vulnerabilities from a surveillance standpoint. As such, our needs assessment revealed that if the entire dengue information flow were to be metaphorically considered as a sand clock, steps 8 and 9 could be characterized as its neck—the very part that dictates the speed and flow of the process. In other words, the entire informational process curves into and out of the PHI–client interaction.

Recently Labrique et al [24] responded to a persistent criticism about the excessive tendency to report findings from mHealth pilot studies [25], sometimes referred to as "pilotitis," by highlighting emergent, more scientifically robust, mHealth evidence. Our study is uniquely positioned to contribute to this discussion as we generate empirical evidence about the informational and technological needs of health workers in the context of an mHealth intervention. With respect to the specific concern about piloting, our project, through a collaborative, iterative process of software development, was able to effectively complement the traditional pilot approach, and ensure that all 55 PHIs in the CMC (the entire PHI workforce) gradually adopted the system concurrently. It is important to note here that the size of the PHI workforce, the urban nature of Colombo's terrain, and the cooperative nature of CMC's management were factors that facilitated the scale adoption of Mo-Buzz; however, this approach might be less feasible in other public health contexts such as that of the Accredited Social Health Activists in rural India where mHealth innovations need to be introduced in concert with a hierarchical, bureaucratic public health system that is spread out over a vast geographical region. Another focus of Mo-Buzz was to initiate it into the work lives of PHIs in a manner that would cause least disruption to the flow of their daily activities. Although we partially achieved this goal, concerns surrounding the simultaneous management of both, the old paper-based system and the new tablet-based system (Mo-Buzz) continue to exist. These issues will likely be completely ironed out once the paper-based system is fully replaced by a digital version. Similarly, although we continue to work with CMC to formalize the digital reports sent by the PHIs through the tablets, the existing conventions demand added manual tasks, such as a hand-written signature on all completed forms. As we strategize a gradual transfer of technology to CMC, it remains to be seen whether organizational will can be matched with support in terms of funding and technical expertise to streamline organizational processes. Finally, latest reports show large spikes in PHI adoption of Mo-Buzz, to overcome traditional fidelity to paper-based logbooks and handwritten signatures. We expect that over time, more PHIs will be habituated to Mo-Buzz. Contrastingly, we were encouraged to note from some of the PHIs that they were beginning to use the tablets for personal use and other work tasks apart from dengue, such as capturing and storing pictures, sending emails, and so forth. We anticipate that these habits will bear spillover effects for specific use of Mo-Buzz for dengue-related tasks and increase its effective adoption in future.

## Conclusions and Future Work

The Mo-Buzz intervention was first soft launched in June 2013 among a small group of PHIs chosen by the CMC management following which the project team worked on multiple iterations over the next year and half. The system has been fully adopted by the CMC in early 2015 for use by all PHIs. Even so, we expect a gradual adoption curve given the entrenched nature of existing systems in PHI's work habits. Equally, we also expect variable adoption of the system with respect to its specific components and specific PHI subgroups (younger vs older).

The Mo-Buzz intervention is a response to multiple calls by researchers and the policymaking community for collaborations in the area of mobile interventions for global public health. Our experience revealed that the benefits of this paradigm lies in alleviating country-specific public health challenges through a commonly shared understanding of cultural and ethnic mores and sharing of knowledge and technologies. In the next phase, the research team plans to conduct a theoretically informed, mixed-methods evaluation to assess adoption effectiveness and system performance and its effects on dengue program management metrics of the CMC. We also plan to quantitatively compare the user experiences of the Mo-Buzz system with its paper-based predecessor. We call upon future researchers to further dissect the applicability of the Spiral Model of software development to mHealth interventions and contribute to the mHealth evidence debate from theoretical and applied perspectives.

## Acknowledgments

This research is supported by the National Research Foundation, Prime Minister's Office, Singapore under its International Research Centres in Singapore Funding Initiative and administered by the Interactive Digital Media Programme Office. The authors would also like to thank Mr Karthikayen Jayasundar for assistance with coding. This study would not have been possible without the cooperation of the Colombo Municipal Council management, the Public Health Inspectors, Mobitel, and the University of Colombo School of Computing.

## Conflicts of Interest

None declared.



XSL•FO
RenderX

## References


1. Ministry of Defense. Disease Surveillance: Trends. 2015. URL: http://www.epid.gov.lk/web/index.php?option=com_casesanddeaths&Itemid=448&lang=en# [accessed 2016-05-06]
2. Hettiarachchi K, Jeewandara S, Munasinghe A, Perera S. More dengue victims: Staff shortage, ignorance of people, poor garbage management to blame. 2013. URL: http://www.sundaytimes.lk/130609/news/more-dengue-victims-staff-shortage-ignorance-of-people-poor-garbage-management-to-blame-47982.html [accessed 2015-09-09] [WebCite Cache ID 6hmz7eldG]
3. Gopalan S, Mohanty S, Das A. Assessing community health workers' performance motivation: a mixed-methods approach on India's Accredited Social Health Activists (ASHA) programme. BMJ Open 2012 Sep;2(5):e001557 [FREE Full text] [doi: 10.1136/bmjopen-2012-001557] [Medline: 23019208]
4. Bajpai N, Dholakia R. Globalcenters. 2011. Improving the performance of accredited social health activists in India URL: http://globalcenters.columbia.edu/files/cgc/pictures/Improving_the_Performance_of_ASHAs_in_India_CGCSA_Working_Paper_1.pdf [accessed 2015-05-26] [WebCite Cache ID 6hmzWgbEC]
5. Dieleman M, Cuong PV, Anh LV, Martineau T. Identifying factors for job motivation of rural health workers in North Viet Nam. Hum Resour Health 2003 Nov 5;1(1):10 [FREE Full text] [doi: 10.1186/1478-4491-1-10] [Medline: 14613527]
6. Buykx P, Humphreys J, Wakerman J, Pashen D. Systematic review of effective retention incentives for health workers in rural and remote areas: towards evidence-based policy. Aust J Rural Health 2010 Jun;18(3):102-109. [doi: 10.1111/j.1440-1584.2010.01139.x] [Medline: 20579020]
7. McCoy D, Bennett S, Witter S, Pond B, Baker B, Gow J, et al. Salaries and incomes of health workers in sub-Saharan Africa. Lancet 2008 Feb 23;371(9613):675-681. [doi: 10.1016/S0140-6736(08)60306-2] [Medline: 18295025]
8. Willis-Shattuck M, Bidwell P, Thomas S, Wyness L, Blaauw D, Ditlopo P. Motivation and retention of health workers in developing countries: a systematic review. BMC Health Serv Res 2008;8:247 [FREE Full text] [doi: 10.1186/1472-6963-8-247] [Medline: 19055827]
9. International Telecomunication Union. 2013. ITU statistical market overview: Sri Lanka URL: http://www.itu.int/net/newsroom/GSR/2012/reports/stats_sri_lanka.aspx [accessed 2015-05-26] [WebCite Cache ID 6hmzsfWdQ]
10. GSMA Intelligence. 2013. Country overview: Sri Lanka URL: https://gsmaintelligence.com/research/?file=131003-sri-lanka.pdf&download [accessed 2013-05-11]
11. Chang AY, Parrales ME, Jimenez J, Sobieszczyk ME, Hammer SM, Copenhaver DJ, et al. Combining Google Earth and GIS mapping technologies in a dengue surveillance system for developing countries. Int J Health Geogr 2009;8:49 [FREE Full text] [doi: 10.1186/1476-072X-8-49] [Medline: 19627614]
12. 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]
13. Ditsuwan T, Liabsuetrakul T, Chongsuvivatwong V, Thammapalo S, McNeil E. Assessing the spreading patterns of dengue infection and chikungunya fever outbreaks in lower southern Thailand using a geographic information system. Ann Epidemiol 2011 Apr;21(4):253-261. [doi: 10.1016/j.annepidem.2010.12.002] [Medline: 21376272]
14. Hernández-Ávila JE, Rodríguez M, Santos-Luna R, Sánchez-Castañeda V, Román-Pérez S, Ríos-Salgado VH, et al. Nation-wide, web-based, geographic information system for the integrated surveillance and control of dengue fever in Mexico. PLoS One 2013;8(8):e70231 [FREE Full text] [doi: 10.1371/journal.pone.0070231] [Medline: 23936394]
15. Heintze C, Velasco GM, Kroeger A. What do community-based dengue control programmes achieve? A systematic review of published evaluations. Trans R Soc Trop Med Hyg 2007 Apr;101(4):317-325. [doi: 10.1016/j.trstmh.2006.08.007] [Medline: 17084427]
16. McNaughton D. The importance of long-term social research in enabling participation and developing engagement strategies for new dengue control technologies. PLoS Negl Trop Dis 2012;6(8):e1785 [FREE Full text] [doi: 10.1371/journal.pntd.0001785] [Medline: 22953011]
17. Curioso WH, Mechael PN. Enhancing 'M-health' with south-to-south collaborations. Health Aff (Millwood) 2010 Feb;29(2):264-267 [FREE Full text] [doi: 10.1377/hlthaff.2009.1057] [Medline: 20348071]
18. Davis F. Perceived usefulness, perceived ease of use, and user acceptance of information technology. MIS Quarterly 1989 Sep;13(3):319-340.
19. Sekhon L, Lee J, Mehta S, Mukherjee T, Sandler B, Copperman AB. Are Euploid Embryos Less Likely to Result in Ectopic Implantation? [16]. Obstet Gynecol 2016 May;127 Suppl 1:6S. [doi: 10.1097/01.AOG.0000483632.74642.05] [Medline: 27176169]
20. Boehm B, Bose P, Horowitz E, Lee M. Software requirements negotiation and renegotiation aids: A theory-W based spiral approach. 1995 May 16 Presented at: 17th International Conference on Software Engineering; April 23-30, 1995; Seattle, WA.
21. Boehm B, Hansen W. The spiral model as a tool for evolutionary acquisition. CrossTalk 2001 Jan;14(5):4-11.




XSL•FO
RenderX

22. Horstick O, Runge-Ranzinger S, Nathan MB, Kroeger A. Dengue vector-control services: how do they work? A systematic literature review and country case studies. Trans R Soc Trop Med Hyg 2010 Jun;104(6):379-386. [doi: 10.1016/j.trstmh.2009.07.027] [Medline: 20400169]
23. Chang MS, Christophel EM, Gopinath D, Abdur RM, Malaria, Vectorborne O, World Health Organization Regional Office, et al. Challenges and future perspective for dengue vector control in the Western Pacific Region. Western Pac Surveill Response J 2011 Apr;2(2):9-16 [FREE Full text] [doi: 10.5365/WPSAR.2010.1.1.012] [Medline: 23908883]
24. Labrique A, Vasudevan L, Chang LW, Mehl G. H_pe for mHealth: more “y” or “o” on the horizon? Int J Med Inform 2013 May;82(5):467-469 [FREE Full text] [doi: 10.1016/j.ijmedinf.2012.11.016] [Medline: 23279850]
25. Tomlinson M, Rotheram-Borus MJ, Swartz L, Tsai AC. Scaling up mHealth: where is the evidence? PLoS Med 2013;10(2):e1001382 [FREE Full text] [doi: 10.1371/journal.pmed.1001382] [Medline: 23424286]

## Abbreviations

**CDF:** Communicable Disease Form
**CMC:** Colombo Municipal Council
**DIF:** Dengue Investigation Form
**EU:** Epidemiological Unit
**GIS:** Geographical Information Systems
**GPS:** Geographical Positioning System
**PHI:** Public Health Inspector

*Edited by G Eysenbach; submitted 12.05.15; peer-reviewed by R Jayasuriya, D Davies; comments to author 31.01.16; revised version received 22.03.16; accepted 20.04.16; published 01.07.16.*

*Please cite as:*
*Lwin MO, Vijaykumar S, Rathnayake VS, Lim G, Panchapakesan C, Foo S, Wijayamuni R, Wimalaratne P, Fernando ONN*
*A Social Media mHealth Solution to Address the Needs of Dengue Prevention and Management in Sri Lanka*
*J Med Internet Res 2016;18(7):e149*
*URL: http://www.jmir.org/2016/7/e149/*
*doi:10.2196/jmir.4657*
*PMID:27369296*

©May O. Lwin, Santosh Vijaykumar, Vajira Sampath Rathnayake, Gentatsu Lim, Chitra Panchapakesan, Schubert Foo, Ruwan Wijayamuni, Prasad Wimalaratne, Owen Noel Newton Fernando. Originally published in the Journal of Medical Internet Research (http://www.jmir.org), 01.07.2016. This is an open-access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/2.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on http://www.jmir.org/, as well as this copyright and license information must be included.



XSL•FO
RenderX

Original Paper

# Linguistic and Cultural Adaptation of a Computer-Based Counseling Program (CARE+ Spanish) to Support HIV Treatment Adherence and Risk Reduction for People Living With HIV/AIDS: A Randomized Controlled Trial

Ann E Kurth[1], CNM, MPH, PhD; Nok Chhun[1], MS, MPH; Charles M Cleland[2], PhD; Michele Crespo-Fierro[2], RN, AACRN, MS/MPH; José A Parés-Avila[3], RN, MA, ANP-C, DNP, AAHIVS; John A Lizcano[1], MPH; Robert G Norman[4†], PhD; Michele G Shedlin[2], PhD; Barbara E Johnston[5], MD; Victoria L Sharp[6], MD

[1]Yale University School of Nursing, Orange, CT, United States

[2]New York University Rory Meyers College of Nursing, New York, NY, United States

[3]University of Arizona College of Nursing, Tucson, AZ, United States

[4]New York University College of Dentistry, New York, NY, United States

[5]Lincoln Community Health Center, Durham, NC, United States

[6]St. Luke's-Roosevelt Hospital Center, Center for Comprehensive Care, New York, NY, United States

[†]deceased

**Corresponding Author:**
Ann E Kurth, CNM, MPH, PhD
Yale University School of Nursing
400 West Campus Drive
Orange, CT, 06477
United States
Phone: 1 203 785 2393
Fax: 1 203 785 3554
Email: ann.kurth@yale.edu

## Abstract

**Background:** Human immunodeficiency virus (HIV) disease in the United States disproportionately affects minorities, including Latinos. Barriers including language are associated with lower antiretroviral therapy (ART) adherence seen among Latinos, yet ART and interventions for clinic visit adherence are rarely developed or delivered in Spanish.

**Objective:** The aim was to adapt a computer-based counseling tool, demonstrated to reduce HIV-1 viral load and sexual risk transmission in a population of English-speaking adults, for use during routine clinical visits for an HIV-positive Spanish-speaking population (CARE+ Spanish); the Technology Acceptance Model (TAM) was the theoretical framework guiding program development.

**Methods:** A longitudinal randomized controlled trial was conducted from June 4, 2010 to March 29, 2012. Participants were recruited from a comprehensive HIV treatment center comprising three clinics in New York City. Eligibility criteria were (1) adults (age ≥18 years), (2) Latino birth or ancestry, (3) speaks Spanish (mono- or multilingual), and (4) on antiretrovirals. Linear and generalized mixed linear effects models were used to analyze primary outcomes, which included ART adherence, sexual transmission risk behaviors, and HIV-1 viral loads. Exit interviews were offered to purposively selected intervention participants to explore cultural acceptability of the tool among participants, and focus groups explored the acceptability and system efficiency issues among clinic providers, using the TAM framework.

**Results:** A total of 494 Spanish-speaking HIV clinic attendees were enrolled and randomly assigned to the intervention (arm A: n=253) or risk assessment-only control (arm B, n=241) group and followed up at 3-month intervals for one year. Gender distribution was 296 (68.4%) male, 110 (25.4%) female, and 10 (2.3%) transgender. By study end, 433 of 494 (87.7%) participants were retained. Although intervention participants had reduced viral loads, increased ART adherence and decreased sexual transmission risk behaviors over time, these findings were not statistically significant. We also conducted 61 qualitative exit interviews with participants and two focus groups with a total of 16 providers.



XSL•FO
RenderX

**Conclusions:** A computer-based counseling tool grounded in the TAM theoretical model and delivered in Spanish was acceptable and feasible to implement in a high-volume HIV clinic setting. It was able to provide evidence-based, linguistically appropriate ART adherence support without requiring additional staff time, bilingual status, or translation services. We found that language preferences and cultural acceptability of a computer-based counseling tool exist on a continuum in our urban Spanish-speaking population. Theoretical frameworks of technology's usefulness for behavioral modification need further exploration in other languages and cultures.

**Trial Registration:** ClinicalTrials.gov NCT01013935; https://clinicaltrials.gov/ct2/show/NCT01013935 (Archived by WebCite at http://www.webcitation.org/6ikaD3MT7)

***(J Med Internet Res 2016;18(7):e195)*** doi:10.2196/jmir.5830

**KEYWORDS**

antiretroviral therapy adherence; computer-based counseling; cultural adaptation; HIV; linguistic adaptation; prevention with positives; Technology Acceptance Model; viral load

## Introduction

Human immunodeficiency virus (HIV) disease in the United States disproportionately affects minorities, including Latinos [1]. Barriers such as language are associated with lower antiretroviral therapy (ART) adherence among Latinos, yet ART and interventions for clinic visit adherence are rarely developed or delivered in Spanish. Although treatment of HIV has advanced tremendously with the development of ARTs, these medication regimens require lifelong adherence to achieve therapeutic goals [2-5].

The computer-based counseling tool known as the Computer Assessment & Rx Education for HIV-positives (CARE+) is an evidence-based intervention for people living with HIV and acquired immune deficiency syndrome (AIDS) (PLHA) [6,7]. The purpose of the program is to support users in achieving medication adherence and reduce their risk of secondary HIV infections (also known as "positive prevention"). This program, when evaluated in a university-affiliated public HIV clinic and a community-based AIDS service organization in an English-speaking population in Seattle, was found to be effective in reducing HIV-1 viral load and sexual transmission risk behaviors [6]. Technology tools such as CARE+ present significant opportunities to bridge the gap in health promotion delivery, especially if linguistically and culturally adapted for often-neglected groups such as Latinos. In this paper, we use the term Latino; however, in the literature, Latino and Hispanic are used interchangeably, reflecting a lack of consensus as well as the political and demographic implications of both terms [8].

Latinos make up approximately 17% of the US population, but according to the Centers for Disease Control and Prevention, account for 23% of all new HIV infections reported in 2013 in the United States [9]. Furthermore, Latino men account for 85% of all new infections among Latinos in the United States and 81% acquired HIV infection through sexual contact with another male [9]. In contrast, in 2013, Latino women accounted for 15% of all new infections among Latinos in the United States [9]. It is estimated that one in 36 Latino men and one in 106 Latino women will be diagnosed with HIV at some point in their lives [1].

In this manuscript, we describe the adaptation of the CARE+ tool for a Spanish-speaking population (CARE+ Spanish; Figure 1). The Technology Acceptance Model (TAM) [10] was the conceptual framework that guided the Spanish adaptation of the computer-based counseling program. New information and communication technologies (ICT) must be culturally acceptable if they are to be effective in daily clinical practice, rather than just in the context of a controlled trial. Acceptability is defined as the "degree to which a[n]...intervention or any one of the attributes of the...intervention is perceived by the patient/consumer to be consonant with well-being" [11]. Acceptability centers on perception of an innovation (ie, ICT tool) as clusters of perceived attributes. These can be categorized as follows, using ART as an example: (1) perceived inherent attributes (eg, ART is effective or requires scheduling); (2) perceived associational attributes (eg, ART demands periodic clinic visits; is encouraged by medical personnel, but perhaps not by others); and (3) perceived effects (eg, lifestyle and other changes). Each perceived attribute has cultural meaning, and the individual continually weighs the positive and negative aspects of the attributes throughout treatment. The resulting balancing act influences the acceptability of the intervention. Cultural and linguistic factors determine the perception of relevancy of these attributes, as well as other factors, such as age, life cycle stage, health status, motivation to be healthier, and perceptions of the source of the information and intervention [11,12].

The study aims were to (1) conduct usability testing of CARE+ Spanish; (2) establish a real-world utility of CARE+ Spanish by conducting a 12-month longitudinal randomized controlled trial (RCT) to evaluate the impact of the CARE+ Spanish intervention on outcomes, which included ART adherence, sexual transmission risk behaviors, and HIV-1 viral loads; and (3) to assess technology uptake factors, explore the cultural acceptability of the tool, and perceived technology barriers/facilitators among participants and health care providers using the TAM framework.



XSL•FO
RenderX

**Figure 1.** Screenshots of CARE+ Spanish.




## Methods

The CARE+ program incorporates motivational interviewing and principles of chronic HIV disease self-management to enhance health-promoting behavior. Formative research to test the usability and acceptability of the CARE+ computer-based counseling tool has been previously described [7], and the tool was shown in an English-speaking population to improve ART adherence, viral suppression, and reduced secondary sexual transmission risk behavior [6]. Given that computer-based counseling proved promising in an English-speaking population in Seattle, translation into practice and applicability to other populations that could benefit, such as Spanish-speaking Latinos or Spanish-dominant bilingual Latinos living with HIV, guided the adaptation of the CARE+ tool for a Spanish-speaking population.

The recommended process for adaptation and translation from English to a different language version is the forward-back translation method in order to ensure cultural and linguistic equivalence [13]. Given budgetary constraints, and the large amount of software content, an expert panel review was utilized in place of back translation of the CARE+ tool. The expert panel consisted of clinical experts who were bilingual and bicultural with ancestry from major Latino subgroups (ie, Mexican and Puerto Rican). For the forward translation, a translator with a master's degree experienced with HIV health-related materials



XSL•FO
RenderX

translated the CARE+ content into Spanish. Because the professional translator's work had been used with predominantly Mexican-American populations, a member of our study team (JAP), who is of Puerto Rican origin and has experience in AIDS research with Mexican and Puerto Rican populations, took on the role of assuring quality of translation and applicability across Latino subgroups. Additionally, there was a secondary reviewer, a health educator and curriculum writer with a master of public health, who has done substantial translation work for HIV interventions in California, Florida, Puerto Rico, and the Northeastern United States. She performed final reviews of the translated software content. Spanish-speaking actors of Mexican and Colombian heritage recorded the narration, and text changes were also made if the two voice actors reading the script made recommendations. Then, a panel of bilingual HIV health care providers (two physicians, a nurse practitioner, a nurse, and a health educator) reviewed the translation of the content independently to confirm the appropriateness of the terminology and the minimization of idiomatic regional expressions unique to their culture. After discussion to reach consensus, words identified as idiomatic were deleted and substituted with words that were not specific to only one Latino subgroup. Furthermore, an additional local expert advisory panel (composed of one Spanish-speaking person living with HIV, and HIV providers from medicine, nursing, and social work) was convened in New York City to review the CARE+ Spanish tool content and shorten it for use in a real-world high-volume HIV clinic setting. Once a test-ready version of CARE+ Spanish was available, the program was tested with HIV-positive Spanish-speaking individuals to explore acceptability and usability of the intervention. Feedback from the usability testing was incorporated into the final version of the CARE+ Spanish program used for the RCT. In addition, we conducted exit interviews to explore cultural acceptability of the tool among purposively selected intervention participants, and focus groups among clinic providers to explore program acceptability and system efficiency issues.

## Participants

### *Usability Testing*

Participants for the usability portion of the study were recruited and verbally consented from St Vincent's Catholic Medical Center HIV clinic in New York City the last week of March 2010 (IRB #09-096). Because we were interested in the participants' opinions about CARE+ Spanish, and not its efficacy or effectiveness, participants were informed that they could make up answers and skip questions to avoid disclosing personal health information. Using the "think aloud" method [14,15], participants were observed as they completed the CARE+ Spanish program. The observation was timed and careful notes were taken, paying attention to any difficulties navigating the program or understanding the content. While working with the program, the participant was encouraged to "think aloud" and share thoughts in their own words as they worked through the various tasks. A series of structured questions and verbal prompts were also used to elicit participants' reactions to the program. Participants were given a US $20 MetroCard for their time and to reimburse their transportation costs. As a result of St Vincent's closing in April 2010, the study was moved to St Luke's-Roosevelt Hospital for the initiation of the RCT and remainder of the project.

### *Randomized Controlled Trial*

Study participants for the RCT were recruited in the waiting areas from three urban HIV clinic sites of St Luke's-Roosevelt Hospital in New York City from June 4, 2010 to January 3, 2011. Eligibility criteria were (1) adults (age ≥18 years), (2) Latino birth or ancestry, (3) speaks Spanish (mono- or multilingual), and (4) on antiretrovirals at any of the three clinic study sites.

Written informed consent was obtained from all RCT participants. All study procedures were approved by St Luke's-Roosevelt Hospital, Center for Health Sciences, Institutional Review Board (#10-068) and New York University School of Medicine's Institutional Review Board (#09-0740). This RCT is reported according to the CONSORT checklist [16] and the CONSORT-EHEALTH extension [17] (Multimedia Appendix 1). The RCT participants received a US $20 MetroCard for their time and to reimburse their transportation costs at the end of each study visit (five sessions total and an additional US $20 MetroCard if they participated in the exit interview).

The RCT participants were enrolled by research assistants, some of whom were part of the HIV-positive peer program at St Luke's-Roosevelt Hospital. Participants of the peer program were Spanish-speaking PLHA, who were also receiving care at the St Luke's Roosevelt Hospital HIV clinics. Peers in the program were selected as study staff based on their language skills and experience working with other Spanish-speaking PLHA in the HIV clinics. All research assistants completed human subjects' protection certification, which included Health Insurance Portability and Accountability Act (HIPAA), and received training on the study protocol and procedures, and the use of the CARE+ Spanish program.

### *Focus Groups*

The health care providers who participated in the focus groups were engaged in the care of participants (eg, prescribed antiretrovirals and/or supported ART adherence) and were recruited from the same three urban HIV clinic sites of St Luke's-Roosevelt Hospital in New York City as the RCT participants. The two focus groups took place on February 21, 2012 and February 28, 2012. Participants in the health care focus groups were a variety of psychological and medical professionals who had provided care to PLHA for a wide range of years.

## Intervention

The CARE+ Spanish computer-based counseling program was delivered on touchscreen computers with content based on the following theoretical frameworks: information-motivation-behavior [18], social cognitive role modeling [19], and motivational interviewing [20]; and it was evaluated in a prospective longitudinal two-arm RCT design. Participants were automatically randomized by the software to the control or intervention arms following an anonymous study log in by the user. All participants were guided through the



XSL•FO
RenderX

program with audio narration of all content. The intervention session lasted approximately 45 to 60 minutes; the control session lasted approximately 20 to 30 minutes. The control group received only the computer-based audio-narrated risk assessment, which included questions about sexual risk behaviors, substance use, mental health, social support, partner status and disclosure, ART regimen and adherence in last 7 and 30 days, and side effects. In addition to the computer-based audio-narrated risk assessment, the intervention group received tailored feedback through the skill-building videos, health plan, and printout at the end of the session, within the CARE+ Spanish software program. The skill-building videos automatically launched; afterwards, participants could choose to watch additional videos. Video topics included demonstration of healthy behaviors, such as condom use and medication adherence, and discussions about HIV and provider relationships. In the final step, users developed a risk reduction plan related to either ART adherence or safer sex practices to prevent secondary HIV transmission. After the risk reduction plan, participants could opt to watch more videos. At the conclusion of the session, participants received a printout of their tailored feedback and health promotion plan that they could share with their health care provider.

Both study arms also received standard clinical care per Department of Health and Human Services HIV Guidelines [2]. In addition, the computer-based tool identified participants who were experiencing severe depression as measured by the Patient Health Questionnaire (PHQ-9; score ≥20) [21], intimate partner violence, or suicidal ideation. Per study protocol, case managers were then notified for appropriate follow-up and referral. Each group underwent five sessions total at 3-month intervals (at 0, 3, 6, 9, and 12 months). At the 12-month session, the control group was switched over to the intervention condition. Study sessions were scheduled to coincide with clinic visits whenever possible.

### Exit Interviews

The purpose of the qualitative exit interview (Multimedia Appendix 2) was to explore the cultural acceptability of the tool and assess perceived technology barriers and facilitators among participants. Exit interviews were offered to purposively selected intervention participants using systematic sampling. At the 12-month session, every third eligible candidate was offered a face-to-face exit interview. We purposively sampled females and males, older (age ≥35 years) and younger (age ≤35 years), and US- or foreign-born participants. Exit interviews were performed using a semistructured interview guide by research assistants in either English or Spanish as requested by the participant. The interview guide was developed to explore the main concepts of TAM and other cultural factors known to impact the Latino community, such as stigma related to HIV, language, health insurance coverage, and immigration status. Notes were taken by research assistants in English and/or Spanish, including verbatim quotes to capture illustrative comments from respondents.

### Health Care Provider Focus Groups

Health care providers who participated in the focus groups gave written informed consent and received US $50 for their time and travel. Two focus groups were conducted; the sessions lasted approximately 2 hours. The focus groups were conducted using a semistructured interview guide (Multimedia Appendix 3), which outlined topics to be discussed along with suggested probes. Topics included challenges that providers faced in delivering care to HIV-positive patients, adherence-related issues, and usefulness of the CARE+ Spanish computer-based counseling tool. To improve attendance of the providers, the focus groups were conducted before required staff meetings. Prior to the start of the focus group, the providers were given an opportunity to view the counseling program (tablet and headphones), as well as a sample session printout. Due to scheduling issues, different teams of study staff conducted the two focus groups.

### Outcome Measures

The primary outcome was HIV-1 viral load collected from medical chart reviews. This biomarker, along with adherence to medications (measured by 30-day visual analog scale [VAS]) and sexual transmission risk behaviors (defined as lack of condom use with either a main or other partner) identified through the CARE+ Spanish program, were the outcome measures for the assessment of intervention effectiveness. These outcome measures were collected at 0, 3, 6, 9, and 12 months. Although the outcome measures were not collected from every participant at each time interval, a minimum of three data points were collected from all participants.

### Sample Size Determination

For the usability testing, sample sizes greater than five participants have been shown to have sufficient power to detect the majority of usability problems [22]. For the RCT, sample size was calculated based on the target intervention effect on the proportion of participants who are ART adherent, HIV viral load at log10 scale, and occurrence of unprotected sex with HIV-negative/unknown partner(s). All calculations control type I error rate at 0.05. Considering a decrease of 0.5 log10 HIV viral load as a meaningful reduction, with 200 retained participants in each group in a time point-specific post hoc test, there was an expected >97% power to detect this difference with a standard deviation of 1.25 (effect size=0.4).

### Statistical Analysis

Fisher exact and Wilcoxon rank sum tests assessed differences between intervention and control groups in population study characteristics at the baseline assessment. Linear and generalized mixed linear effect models were used to longitudinally assess differences between the intervention and control groups at all available time points for sexual transmission risk behavior, medication adherence (30-day VAS), and viral load variables. These models accommodate missing data (equal numbers of measurements and time intervals between measurements were not required) and, therefore, do not require deletion of participants with incomplete data. A $P \leq .05$ was used as the cutoff for significance. Primary outcomes included ART adherence, sexual transmission risk, and HIV-1 viral loads. Baseline analyses were performed using SAS version 9.3 (SAS Institute Inc, Cary, NC, USA) and, for the longitudinal analysis, the lme4 package [23] of the R statistical computing



XSL•FO
RenderX

environment [24]. Given that the control group received the intervention at 12 months, data analysis was limited to four time periods (0, 3, 6, and 9 months). Effect sizes for undetectable viral load and sexual transmission risk were presented in the odds ratio metric. For viral load and 30-day VAS, Cohen's $f^2$ was calculated to convey effect sizes for group differences in change over time [25].

### Interview Analysis

Data from the exit interviews were transcribed onto spreadsheets by two researchers from the study team (MGS and MCF), while the focus groups were recorded by a stenographer, with transcripts provided from the two sessions. Data were analyzed using content analysis within a framework of technology transfer [10] to identify factors affecting acceptability, utilization, and impact. The exit interview spreadsheets and the focus group transcripts were analyzed by MS and MCF; emergent themes and issues were categorized by each. Saturation of themes was determined after 61 participant exit interviews and the two provider focus groups. Inconsistencies in the themes were discussed between MS and MCF until consensus was reached; selected quotations were agreed on as salient examples of themes.

## Results

### Usability Testing

Software usability testing was conducted with eight Spanish-speaking PLHA (6 male, 2 female). Five of six males identified their language preference as bilingual and one as English-dominant; one female identified as bilingual and the other as English-dominant. All usability participants reported the program was easy to use and navigate; questions were clear, specific, and understandable. All participants who identified as bilingual (6/8) reported that the Spanish used was basic and easy to understand. The two English-dominant participants reported there were some "big (high-register) words," but they were able to navigate the program and follow instructions without any major problems. All participants agreed that the counseling tool supported privacy and confidentiality, especially for people who are more quiet and reserved about their HIV status. They agreed when using this counseling tool, one can be more open and honest about responses because of the feeling of not being judged. Seven of eight reported that they would prefer to use this tool rather than counseling with a person. Overall, the counseling tool met with everyone's expectations; on average, it was rated a nine out of 10. It was described as user friendly and self-explanatory.

### Randomized Controlled Trial

We approached 1224 individuals at three study sites; 556 consented (45.42% acceptance), 494 were randomized and completed baseline assessments, and 86.2% (426/494) were retained for the 12-month study duration (Figure 2).

Table 1 illustrates participant characteristics at baseline by study arm. There were no significant differences between the treatment and control groups at baseline except that the CARE+ intervention group had a higher proportion of transgendered individuals (4.0%, 9/225 vs 0.50%, 1/206; $P$=.05) and were younger than those in the control group (mean 46.8, SD 9.7 vs mean 48.9, SD 9.1; $P$=.02, respectively).

Figure 3 shows the mean and 95% confidence intervals for outcome means or proportions by time point and treatment condition. Although intervention participants had reduced viral loads, increased ART adherence, and decreased sexual transmission risk behaviors over time, patterns of change in the intervention group were not more favorable than in the control group.

Figure 4 summarizes main outcomes of interest at each follow-up time point. Figure 4 A illustrates 95% confidence intervals for log10 viral load mean differences and change at each different time point as well as overall change in the control versus the CARE+ intervention group. Although there was a decrease in viral loads among participants with a detectable load at baseline (greater in the intervention vs control group), this difference in change was not statistically significant. In addition, among those with detectable viral loads at baseline, the CARE+ intervention group had higher odds of being undetectable at the 9-month follow-up when compared to controls (Figure 4 B), but this difference was not statistically significant. Figure 4 C shows 95% confidence intervals for VAS differences at each follow-up time point; although ART adherence was higher in the CARE+ intervention group vs control in the total sample overall and among participants with detectable viral loads at baseline, no differences in change or at any follow-up point were statistically significant. Finally, although sexual transmission risk behaviors decreased over time, when intervention and control groups were compared, no differences in change or at any follow-up point were statistically significant.

Differences between the CARE+ intervention group and the control group in change on viral load were small both for the total sample ($f^2$=0.0003) and for those with detectable viral load at baseline ($f^2$=0.006). Similarly, group differences in change on VAS were small for both the total sample ($f^2$=0.002) and for those with detectable viral load at baseline ($f^2$=0.005). Figures 4 B and 4D show effect sizes for the sexual transmission risk and undetectable viral load outcomes in the odds ratio metric. Differences in undetectable viral load between the CARE+ intervention group and the control group were consistently small at baseline and across follow-ups. Differences in sexual transmission risk were small at baseline and became even smaller with each follow-up.



XSL•FO
RenderX

**Table 1.** Demographic characteristics of CARE+ Spanish intervention and control groups (N=433).

| Variable | CARE+ Spanish (n=226) | Control (n=207) | *P* [a] |
|---|---|---|---|
| **Sex, n (%)** | | | |
| Male | 153 (68.0) | 143 (69.4) | .05 |
| Female | 56 (24.9) | 54 (26.2) | |
| Transgender | 9 (4.0) | 1 (0.5) | |
| Unknown | 7 (3.1) | 8 (3.8) | |
| **Age (years), mean (SD)** | 46.8 (9.7) | 48.9 (9.1) | .02 |
| **Ethnicity, n (%)** | | | |
| Latino | 218 (96.9) | 191 (92.7) | .20 |
| **Race, n (%)** | | | |
| American Indian or Alaska Native | 12 (5.3) | 13 (6.4) | .33 |
| Asian | 2 (0.8) | 0 (0.0) | |
| Black or African American | 13 (5.8) | 19 (9.2) | |
| Native Hawaiian or other Pacific Islander | 2 (0.8) | 3 (1.5) | |
| White | 38 (16.8) | 42 (20.4) | |
| Other race | 151 (66.8) | 118 (57.3) | |
| Multiple race | 3 (1.3) | 5 (2.4) | |
| Unknown | 5 (2.2) | 7 (3.4) | |
| **Education, n (%)** | | | |
| No high school diploma/GED | 79 (35.0) | 72 (34.8) | .13 |
| High school diploma/GED only | 65 (28.8) | 75 (36.2) | |
| More than high school | 75 (33.2) | 56 (27.1) | |
| Unknown | 7 (3.1) | 4 (1.9) | |
| **Substance use behavior, n (%)** | | | |
| Ever injecting drug use | 12 (5.3) | 10 (4.8) | .55 |
| Alcohol abuse | 44 (19.6) | 41 (19.8) | >.99 |
| Methamphetamine use | 12 (5.5) | 12 (6.1) | .84 |
| Crack/cocaine use | 24 (11.2) | 25 (12.6) | .76 |
| Intimate partner violence, n (%) | 16 (7.2) | — | — |
| **Sexual behavior, n (%)** | | | |
| Any sex past 3 months | 139 (63.8) | 124 (62.3) | .76 |
| Risky sex[b] | 35 (25.9) | 26 (21.7) | .46 |
| Condom use with problems | 47 (33.8) | 36 (29.0) | .43 |
| Any sex without condoms or with condom problems | 69 (50.7) | 56 (46.3) | .53 |
| Discordant sex with main partner | 8 (3.7) | 7 (3.5) | >.99 |
| Discordant sex with other partner | 11 (5.1) | 6 (3.1) | .46 |
| **ART Adherence** | | | |
| Adherence VAS, mean (SD) | 87.0 (22.4) | 89.6 (19.5) | .20 |
| VAS Scale ≥95%, n (%) | 135 (59.7) | 122 (58.9) | .92 |
| Missed doses, mean (SD) | 1.5 (8.0) | 1.6 (8.2) | .86 |
| 1 or more missed doses past 7 days, n (%) | 74 (37.2) | 64 (35.2) | .75 |
| 2 or more missed doses past 7 days, n (%) | 41 (20.6) | 36 (19.8) | .90 |



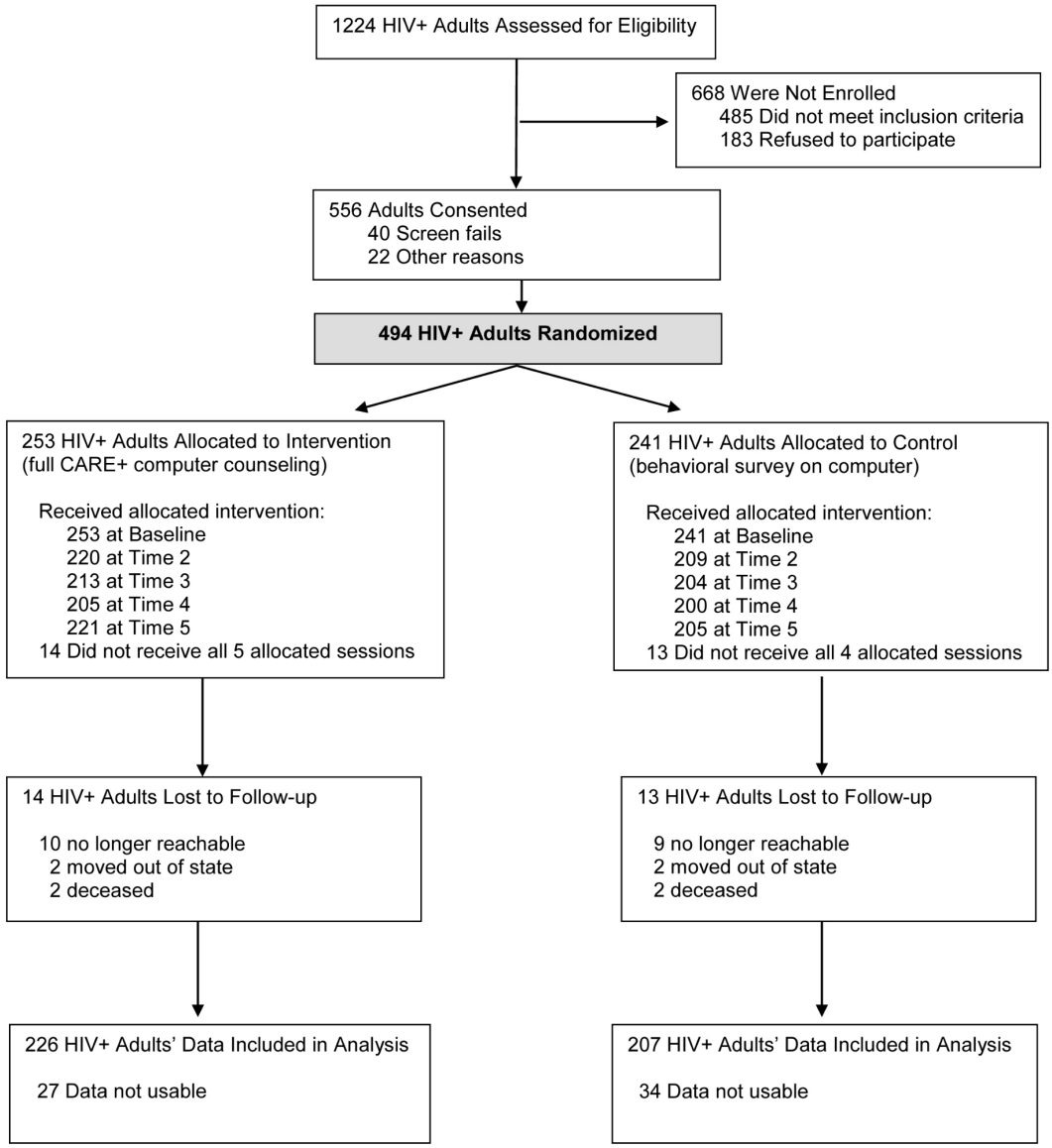
XSL•FO
RenderX

| Variable | CARE+ Spanish (n=226) | Control (n=207) | *P* [a] |
|---|---|---|---|
| > VAS 95% + no missed doses, n (%) | 93 (44.1) | 89 (44.9) | .92 |
| Log10 HIV-1 viral load, mean (SD) | 1.2 (1.8) | 1.2 (1.6) | .93 |
| Detectable viral load, n (%) | 78 (34.5) | 76 (36.7) | .69 |
| Ever told resistant virus, n (%) | 30 (13.8) | 32 (15.9) | .67 |
| Years since HIV diagnosis, mean (SD) | 12.6 (6.9) | 13.4 (7.3) | .23 |
| Depression (PHQ-9), n (%) | 35 (15.6) | 39 (18.8) | .38 |

[a] Comparisons by Fisher exact test or Wilcoxon rank sum test; All categorical tests performed without the unknown category.

[b] Did not use a condom with either main partner or other partner.

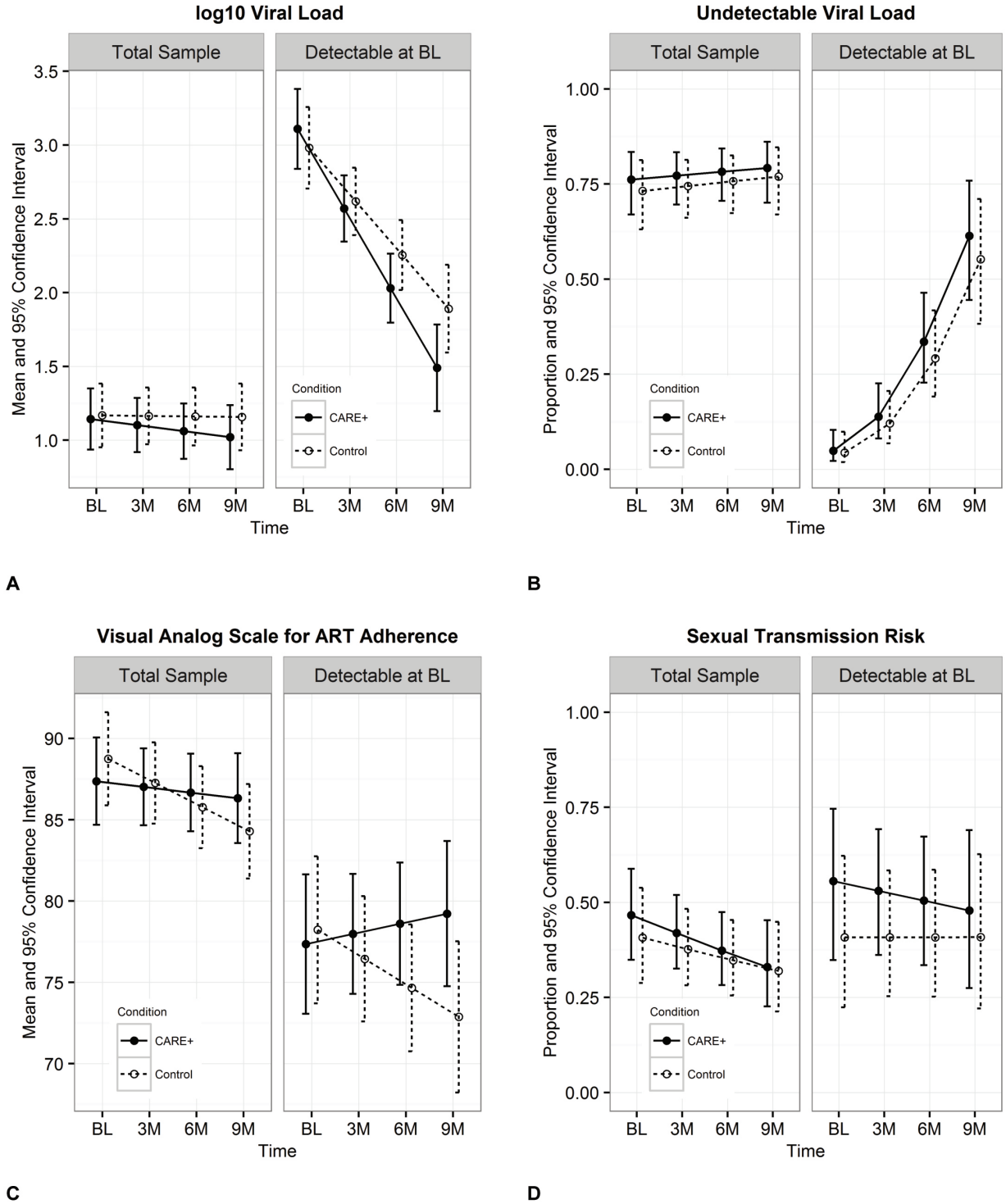
**Figure 2.** Participant flowchart of the CARE+ Spanish computer-based counseling intervention trial, five sessions over 12 months.






XSL•FO
RenderX

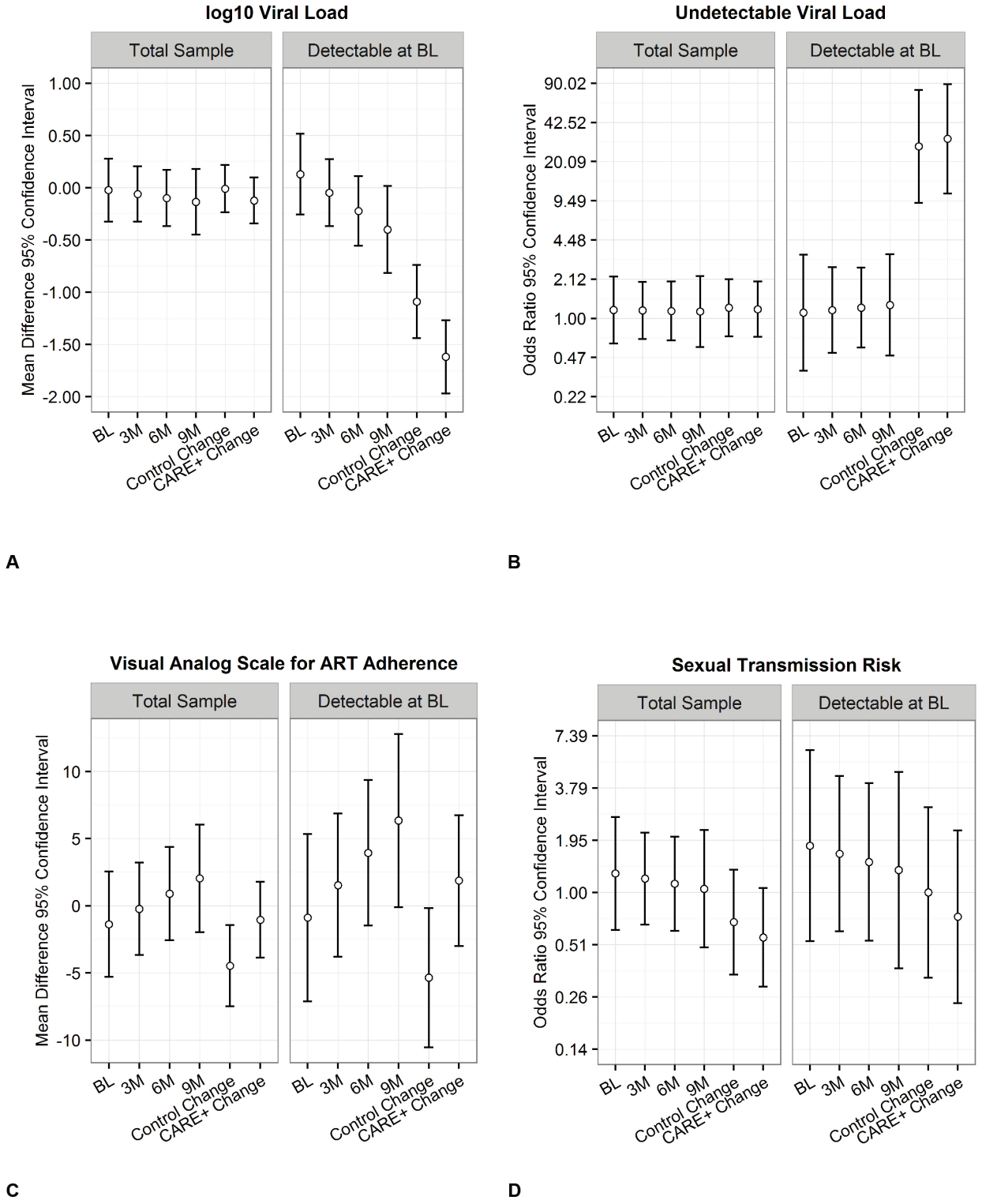
**Figure 3.** Adjusted mean values or proportions by time and treatment condition of the total sample and those with detectable viral loads at baseline (BL) for (a) log10 viral load, (b) undetectable viral load, (c) visual analog scale for ART adherence, and (d) sexual transmission risk. Whiskers represent 95% confidence intervals.






XSL•FO
RenderX

**Figure 4.** Mean differences or odds ratios contrasting CARE+ and control conditions at each follow-up time point, and baseline (BL) and final time points within each condition, for (a) log10 viral load, (b) undetectable viral load, (c) visual analog scale for ART adherence, and (d) sexual transmission risk. Whiskers represent 95% confidence intervals.




## Exit Interviews

The open-ended exit interview identified a range of HIV-related concerns, lack of confidence in providers, multiple sources for HIV-related information, and experiences using the CARE+ Spanish computer-based counseling program.

Characteristics of exit interview participants (N=61) are summarized. In all, 37 (61%) men, 21 (34%) women, and 3 (5%) transgendered persons participated in the interviews. Participant ages ranged from 21 to 69 years, with a mean age of 48.0 (SD 12.0) years. The participants included those born in the United States and Puerto Rico, and immigrants from



XSL•FO
RenderX

Colombia, Costa Rica, the Dominican Republic, Guatemala, Honduras, Mexico, Nicaragua, and Panama, with a range of 4 to 50 years residing in the United States.

### *Concerns Related to Human Immunodeficiency Virus*

In total, 95% (58/61) had disclosed their HIV status to someone. The main issues affecting the interviewed participants included insurance/benefits-related issues, side effects from medications (both antiretrovirals and other medications), being able to work, immigration documentation, and housing issues. The majority of participants did not feel that their issues related to medications or the virus were any different because they were Latino. However, some did feel there was a difference if the issue(s) related to documentation (ie, "having papers"). One participant noted, "illegal status, it leaves us without options." Another participant shared additional comments regarding obstacles faced by their undocumented status: "Of course, because we are immigrants and we don't have legal papers to get around or anything." Some noted that English-speaking/non-Latinos have more information and receive different care. Other participants discussed the problems Latinos face that keep them from focusing on their illness, fears of engaging in care, not trusting to reveal their diagnosis, parents not teaching their children about safer sex, and fears relating to stigma and discrimination: "Spanish people appear stronger around other(s) and don't share their weakness" and "Latinos have machista behavior in not taking care of themselves..." Language barriers were a concern in accessing information, understanding, and being understood by providers and navigating the health care system.

### *Multiple Sources of Information Related to Human Immunodeficiency Virus*

Participants reported a range of resources for obtaining HIV-related and general health information. Overwhelmingly, doctors and clinic staff were reported to be reliable sources of information because they were said to have studied and practiced HIV care for some time: "I put my health in her hands, because she knows what is good and what isn't for me." (Note that participants may be referring to their prescribing providers as "doctors" whether they are medical doctors, nurse practitioners, or physician assistants.)

Some participants purposefully seek out other sources for HIV and general health information. The Internet, because "broader information (is) available," and the Centers for Disease Control and Prevention "made the first concrete studies on managing care" were more trusted sources for some clients. Others cited support groups and peers at AIDS service organizations and clinics as expert voices for comparison of experiences. The media (radio, television, and print) were identified as potential sources of information for the general public. Lastly, families, friends, and spiritual leaders were seen as another source of information for some.

A majority of participants reported that they did receive messages from their providers regarding the importance of taking their medications regularly, having a sexual life with HIV/AIDS, and measures to prevent the transmission of HIV to others. Some participants reported not receiving medication information from their providers: "They have never told me anything." Others knew that medications lessen viral loads, increase $CD_4$ counts, avoid creating resistance, reduce infections, and maintain overall health. One participant added, "It is a marriage with the medications." Some participants were motivated to have these discussions about taking the medications because they reported needing to be there for their children and to maintain their quality of life.

For some participants, any discussion of sexual behavior was one they were not comfortable having with their providers. A few reported giving up on relationships and sex after receiving their HIV diagnosis. Overall, there was a clear awareness that condoms should be used to protect themselves and their partners from HIV and other sexually transmitted infections. Communication was discussed as another way to enhance protection. One of the participants stated that he did not use condoms with his wife who was also HIV-positive. Another discussed using withdrawal when his viral load is undetectable. A few participants stated that their providers did not discuss this topic with them. The participants' answers were similar when discussing modes of reducing transmission (ie, condom use and open discussions with partners). They added abstinence and masturbation as ways to practice safer sex. Medicines were also seen as helping reduce transmission to infants.

For the most part, the majority of participants were comfortable discussing medications and sex with their providers. Some noted that they initiated discussions on medications and sex, whereas for others it was their providers. Some stated that they had "confidence" in their providers, whereas others felt it was necessary for their own well-being to take the initiative: "After receiving treatment with the same doctor for so long, this person becomes part of your family." A salient finding was that some participants did not discuss medications with their providers because they felt they had other resources for information. Regarding sex, some reported preferring to speak with a provider of the same sex or same sexual orientation. A few did not have any discussions regarding medications or sex with their providers: "They don't have the time." Another participant summed up her sexual situation and the need to discuss it as follows: "There's not much to say, my status was so traumatic that I don't believe in love."

### *CARE+ Spanish Computer-Based Counseling Program*

The majority of participants reported a positive experience with using the tablet computer. They thought it was "interesting," "easier to talk to than a person," "it was like having another doctor," "better than having a piece of paper," "the computer doesn't judge," and "I educated myself and came out of such ignorance." On the other hand, some had some issues with the questions being repetitive and the structure of the sentences. Participants reported learning more about HIV, their medications, and things they would not think of asking the provider.

"Likes" of the computer included the ease of use, the touchscreen, the videos, the confidentiality (through the headphones), and the information provided about medications, safer sex, the narrators, and avatars. "Dislikes" of the computer were the long sessions, computer problems such as taking a long time to reboot, the videos, the headphones, not knowing



XSL•FO
RenderX

how to silence the program, and not having an option for English. Some participants were not comfortable with the directness of the language and some of the topics (drugs, sexual abuse). One participant noted, "Some questions were very strong, too direct, and a bit long." The most salient objection, however, was the perceived redundancy of the content.

All the participants reported a sense of privacy and confidentiality while using the computer. Some were concerned about this at the beginning of the study and this was noted to be one of the reasons for some anxiety when first using the computer. Other reasons for anxiety at the start of the study were being unfamiliar with computers and unsure of Spanish language proficiency. Participants spoke of becoming familiar with the format of the program and having the peers to help them get used to it. The availability of the peers was viewed as an overwhelmingly positive aspect of the program.

Although the majority of those interviewed preferred the program in Spanish, some would have preferred to have an English version available and some thought having both languages would work best for them. (This is important to note for any replication in an urban environment where the population is likely to include acculturated and bilingual individuals.) Nearly all participants stated that they would use the program again to pass the time while waiting in the clinic, and when new information could be provided. Some would use it at every visit, whereas most opted for a few times a year. A few participants were clear in not wanting to use the computer program again; one person disliked the voice of the narrator and the other was uncomfortable with the computer. One participant stated that the program might be good for new clients at the clinic or newly diagnosed people.

Overall, the CARE+ Spanish program was viewed positively. Most participants used it to improve their health, learn their medications (eg "Showed me how much I know, tested me on what I knew, allowed me to be honest with myself"), and change some behavior (eg "I loved it! I learned so much and because of this I slowed down my sex life and am more careful"). Although some participants saw the program as an important part of their care by asking questions of their providers, others did not, seeing it as not relevant to their lives (eg "the long explanations of topics that were not relevant to me"). The spectrum of responses received in the exit interview reflects the diversity of the study population, Latinos living in New York City.

### Health Care Provider Focus Groups

The first focus group was conducted on February 21, 2012 with seven participants. Participants included three psychologists, three psychiatric nurse practitioners, and one psychiatrist. Years in health care ranged from 4 to 23 years and years in HIV care ranged from 1.5 to 20 years. There were three bilingual providers present. Providers represented all three study sites; time at their respective sites ranged from 1.5 to 15 years.

The second focus group was conducted on February 28, 2012 with nine participants, but only eight participants provided their demographic information. Participants included five medical doctors and three nurse practitioners. Years in health care ranged from 14 to 29 years and years in HIV care ranged from 11 to 14 years. There were three bilingual providers present. Providers represented all three study sites and time at their respective sites ranged from 5 months to 16 years.

Providers expressed frustration in continuing to confront the same adherence obstacles over time (eg, lack of consistent safer sex practices). However, this was a general finding and not specific to Latino patients. One provider reported that not using condoms makes the patient feel "normal" (eg, "I don't feel sick with HIV when I don't use a condom"). Other adherence obstacles that are in agreement with findings from the RCT exit interviews include substance abuse, lack of documentation, stigma, and trauma. According to one provider, trauma is a "huge, huge factor" because it affects people, their belief in their right to protect themselves, and their expectations that intimate encounters are at least safe for both parties. Overall, providers were in agreement that "like a hydra head," when one issue gets addressed, another comes up.

Regarding the usefulness of the CARE+ Spanish tool, providers agreed that multiple approaches are important (eg "more education is always good") and that patients may feel more empowered to have a resource (eg "an education tool" and "a health enhancement tool") they can access without help. However, they felt that there are no substitutes for the provider-patient relationship, personalism in Latino culture, and loyalty to a provider: "...technology might not have the same impact." Although intervention participants had the option to share their health promotion plan printout and discuss their study participation with their provider, the providers reported that none of the participants did. Therefore, a significant limitation on provider feedback about the CARE+ Spanish program was their lack of familiarity with the tool.

## Discussion

### Principal Results

The CARE+ computer-based counseling tool adapted for a Spanish-speaking population (CARE+ Spanish) was acceptable and feasible to implement in an urban clinic setting. Participants liked the ease of use and the sense of privacy and confidentiality that the computer-based counseling tool provided. The health care providers agreed that multiple approaches are needed and that the counseling tool can be an additional resource for HIV care and support. The CARE+ Spanish program demonstrated trends in positive impact in reducing viral loads, increasing ART adherence and decreasing risky sexual behaviors in three comprehensive care clinics in New York City. In the CARE+ Spanish trial, differences between arms were not statistically significant. In contrast, the CARE+ computer-based counseling tool, when evaluated in a university-affiliated public HIV clinic and a community-based AIDS service organization in Seattle, was found to be efficacious in reducing HIV-1 viral load and sexual transmission risk behaviors [6]. These two randomized clinical trials highlight the importance of targeting the right populations when adapting technology tools to support patient treatment engagement. Although other computer-based behavioral interventions delivered in a clinic setting have been found to improve self-reported antiretroviral adherence [26]



XSL•FO
RenderX

and reduce risky sexual behaviors [27,28], to the best of our knowledge, CARE+ Spanish is the first non-English language computer-based counseling program to provide medication adherence support and promote positive prevention in a HIV-positive minority population. Another version of the counseling tool, CARE+ Kenya, linguistically and culturally adapted for use in clinic settings in Kenya (ClinicalTrials.gov: NCT01015989), may provide additional insights regarding the efficacy of computer-based counseling interventions across different populations.

### Limitations and Strengths

We draw from the qualitative data to provide potential explanation of the statistical lack of effect noted from this intervention in this population and setting. Exit interviews from intervention participants and focus groups with providers highlight the efficacy of computer-based counseling tools in overcoming adherence challenges experienced by culturally and linguistically diverse communities, especially stigma. A study that explored an intervention to engage PLHA to initiate ART, found that by sampling participants primarily from a clinic setting, they were encountering individuals who had already overcome many of the barriers to initiating and adhering to ARTs [29]. It is possible, although this was not explored, that our participants were also further along in the HIV treatment cascade [30], and this may have been a factor in reducing the effect of the CARE+ Spanish intervention. The computer-based audio-narrated risk assessment that both groups received at baseline may have been enough to support the maintenance of adherence in both groups, and the positive trend noted in the treatment group explained by the impact of the full intervention. Additionally, health provider acceptability is important for any ICT tool that aims to be incorporated into real-world practice. Therefore, although the CARE+ Spanish intervention was developed as a stand-alone computer-based counseling tool, participants did receive a printout of their session that they could use to initiate conversation with their provider about adherence support and risk reduction. Instead, providers were unfamiliar with the CARE+ Spanish program and reported that participants did not share the health plan printouts with them. Although not a specific outcome measure, the lack of familiarity with the specifics of the program and the session printout, from the providers, may highlight that the participants were not engaging their providers on this aspect of their self-care, and this may have been a factor in reducing the effect of the intervention. Potential improvement of this counseling intervention may be developing a mobile cloud-based platform to support users in self-motivated behavioral change for better health. Additionally, targeting the intervention to individuals with adherence and/or sexual transmission risk problems may be another way of improving the program's effectiveness.

An innovative strength of our study is the use of peers to support ICT use in a clinic setting for an important and often-neglected population that is disproportionately affected by HIV disease burden. Peer involvement in programs designed to impact attitudes and behaviors have been shown to be effective [31,32].

### Conclusion

A computer-based counseling tool grounded in the TAM theoretical model and delivered in Spanish was acceptable and feasible to implement in a high-volume HIV clinic setting. It can provide evidence-based, linguistically appropriate ART adherence support without requiring additional staff time, bilingual status, or translation services. We found that language preferences and cultural acceptability of a computer-based counseling tool exist on a continuum in our urban Spanish-speaking population. Theoretical frameworks of technology's usefulness in behavioral modification needs further exploration in other languages and cultures to determine where on the HIV care and treatment continuum these interventions may have the greatest impact.

## Acknowledgments

We thank the following deeply for their contributions and support: Jim Larkin, David Larkin, Carter Wickstrom, David Thomas, and Andrew Choi of Resources Online (www.ronline.com/care); Danielle Beiling and Zachariah Hennessey at St Luke's-Roosevelt Hospital; staff members at the Center for Comprehensive Care; our research assistants; and the study participants.

The project described was supported by Award Number RC1MH088307 from the National Institute of Mental Health. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institute of Mental Health or the National Institutes of Health. We thank our project officer Cynthia Grossman. Initial adaptation of CARE+ computer-based counseling tool for a Spanish-speaking population was made possible by a grant to Ann Kurth (principal investigator) and José Parés-Avila (co-investigator) from the University of Washington's Institute of Translational Health Sciences (ITHS). ITHS is funded by Grant Number 1UL1RR025014-01 from the National Center for Research Resources (NCRR), a component of the National Institutes of Health (NIH) and NIH Roadmap for Medical Research.

## Authors' Contributions

AEK conceived of the study and study design. AEK, NC, JAP, MCF, JAL, and CMC drafted the manuscript. MCF, MGS, RGN, and CMC were responsible for data analysis. JAP conducted the formative research that guided the adaptation of the CARE+ tool for a Spanish-speaking population. JAL managed the study. All authors read and approved the final manuscript.

## Conflicts of Interest

None declared.



XSL•FO
RenderX

### Multimedia Appendix 1

Consort-Ehealth Checklist.

[PDF File (Adobe PDF File), 630KB - jmir_v18i7e195_app1.pdf ]

### Multimedia Appendix 2

CARE+ Spanish Participant Interview.

[PDF File (Adobe PDF File), 484KB - jmir_v18i7e195_app2.pdf ]

### Multimedia Appendix 3

CARE+ Spanish Topic Guide: Clinic Staff Focus Group.

[PDF File (Adobe PDF File), 437KB - jmir_v18i7e195_app3.pdf ]

## References


1. Centers for Disease Control and Prevention. Fact sheet: HIV among Hispanics/Latinos in the United States and dependent areas. 2013 Nov. URL: http://www.cdc.gov/hiv/pdf/risk_latino.pdf [accessed 2016-03-30] [WebCite Cache ID 6gP5vqXwU]
2. Panel on Antiretroviral Guidelines for Adults and Adolescents. Guidelines for the Use of Antiretroviral Agents in HIV-1-Infected Adults and Adolescents. 2016 Jan 28. URL: https://aidsinfo.nih.gov/contentfiles/lvguidelines/adultandadolescentgl.pdf [accessed 2016-03-30] [WebCite Cache ID 6gP6RTnoE]
3. World Health Organization. Adherence to Long Term Therapies: Evidence for Action. Geneva: World Health Organization; 2003. URL: http://www.who.int/chp/knowledge/publications/adherence_full_report.pdf [accessed 2016-03-30] [WebCite Cache ID 6gP6a4rK7]
4. Simoni JM, Pearson CR, Pantalone DW, Marks G, Crepaz N. Efficacy of interventions in improving highly active antiretroviral therapy adherence and HIV-1 RNA viral load. A meta-analytic review of randomized controlled trials. J Acquir Immune Defic Syndr 2006 Dec 1;43 Suppl 1:S23-S35 [FREE Full text] [doi: 10.1097/01.qai.0000248342.05438.52] [Medline: 17133201]
5. Rueda S, Park-Wyllie LY, Bayoumi AM, Tynan AM, Antoniou TA, Rourke SB, et al. Patient support and education for promoting adherence to highly active antiretroviral therapy for HIV/AIDS. Cochrane Database Syst Rev 2006(3):CD001442. [doi: 10.1002/14651858.CD001442.pub2] [Medline: 16855968]
6. Kurth AE, Spielberg F, Cleland CM, Lambdin B, Bangsberg DR, Frick PA, et al. Computerized counseling reduces HIV-1 viral load and sexual transmission risk: findings from a randomized controlled trial. J Acquir Immune Defic Syndr 2014 Apr 15;65(5):611-620 [FREE Full text] [doi: 10.1097/QAI.0000000000000100] [Medline: 24384803]
7. Skeels MM, Kurth A, Clausen M, Severynen A, Garcia-Smith H. CARE+ user study: usability and attitudes towards a tablet pc computer counseling tool for HIV+ men and women. AMIA Annu Symp Proc 2006:729-733 [FREE Full text] [Medline: 17238437]
8. Hayes-Bautista DE, Chapa J. Latino terminology: conceptual bases for standardized terminology. Am J Public Health 1987 Jan;77(1):61-68. [Medline: 3789240]
9. Centers for Disease Control. 2013. HIV among Hispanics/Latinos URL: http://www.cdc.gov/hiv/group/racialethnic/hispaniclatinos [accessed 2016-03-30] [WebCite Cache ID 6gP6igE5k]
10. Lai T, Larson EL, Rockoff ML, Bakken S. User acceptance of HIV TIDES--Tailored Interventions for Management of Depressive Symptoms in persons living with HIV/AIDS. J Am Med Inform Assoc 2008;15(2):217-226 [FREE Full text] [doi: 10.1197/jamia.M2481] [Medline: 18096915]
11. Polgar S, Marshall J. The search for culturally acceptable fertility regulating methods. In: Marshall J, Polgar S, editors. Culture, Natality and Family Planning (Monograph 21). Chapel Hill, NC: Carolina Population Center, University of North Carolina; 1976:204-218.
12. World Health Organization. Task Force on Acceptability of Fertility Regulating Methods. Doc. ATF-G(2/73). Geneva: World Health Organization; Feb 1973.
13. Vincent D, McEwen MM, Pasvogel A. The validity and reliability of a Spanish version of the summary of diabetes self-care activities questionnaire. Nurs Res 2008;57(2):101-106. [doi: 10.1097/01.NNR.0000313484.18670.ab] [Medline: 18347481]
14. Jaja C, Pares-Avila J, Wolpin S, Berry D. Usability evaluation of the interactive Personal Patient Profile-Prostate decision support system with African American men. J Natl Med Assoc 2010 Apr;102(4):290-297 [FREE Full text] [Medline: 20437736]
15. Fonteyn ME, Kuipers B, Grobe SJ. A description of think aloud method and protocol analysis. Qual Health Res 1993 Nov 01;3(4):430-441. [doi: 10.1177/104973239300300403]
16. Schulz KF, Altman DG, Moher D. CONSORT 2010 statement: updated guidelines for reporting parallel group randomized trials. Ann Intern Med 2010 Jun 1;152(11):726-732. [doi: 10.7326/0003-4819-152-11-201006010-00232] [Medline: 20335313]




XSL•FO
RenderX

17. 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;13(4):e126 [FREE Full text] [doi: 10.2196/jmir.1923] [Medline: 22209829]
18. Fisher JD, Fisher WA, Amico KR, Harman JJ. An information-motivation-behavioral skills model of adherence to antiretroviral therapy. Health Psychol 2006 Jul;25(4):462-473. [doi: 10.1037/0278-6133.25.4.462] [Medline: 16846321]
19. Bandura A. Health promotion by social cognitive means. Health Educ Behav 2004 Apr;31(2):143-164. [doi: 10.1177/1090198104263660] [Medline: 15090118]
20. Tober G. Motivational interviewing: helping people change. Alcohol Alcoholism 2013 Feb 18;48(3):376-377. [doi: 10.1093/alcalc/agt010]
21. Thibault JM, Steiner RW. Efficient identification of adults with depression and dementia. Am Fam Physician 2004 Sep 15;70(6):1101-1110 [FREE Full text] [Medline: 15456119]
22. Faulkner L. Beyond the five-user assumption: benefits of increased sample sizes in usability testing. Behav Res Methods Instrum Comput 2003 Aug;35(3):379-383. [Medline: 14587545]
23. Bates D, Mächler M, Bolker B, Walker S. Fitting linear mixed-effects models using lme4. J Stat Soft 2015;67(1):1-48. [doi: 10.18637/jss.v067.i01]
24. R Development Core Team. R-A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing URL: https://www.r-project.org/ [accessed 2016-03-30] [WebCite Cache ID 6gP7YVTUL]
25. Selya AS, Rose JS, Dierker LC, Hedeker D, Mermelstein RJ. A practical guide to calculating Cohen's f(2), a measure of local effect size, from PROC MIXED. Front Psychol 2012;3:111 [FREE Full text] [doi: 10.3389/fpsyg.2012.00111] [Medline: 22529829]
26. Fisher JD, Amico KR, Fisher WA, Cornman DH, Shuper PA, Trayling C, LifeWindows Team. Computer-based intervention in HIV clinical care setting improves antiretroviral adherence: the LifeWindows Project. AIDS Behav 2011 Nov;15(8):1635-1646. [doi: 10.1007/s10461-011-9926-x] [Medline: 21452051]
27. Grimley DM, Hook EW. A 15-minute interactive, computerized condom use intervention with biological endpoints. Sex Transm Dis 2009 Feb;36(2):73-78. [doi: 10.1097/OLQ.0b013e31818eea81] [Medline: 19125141]
28. Lightfoot M, Rotheram-Borus MJ, Comulada WS, Reddy VS, Duan N. Efficacy of brief interventions in clinical care settings for persons living with HIV. J Acquir Immune Defic Syndr 2010 Mar;53(3):348-356. [doi: 10.1097/QAI.0b013e3181c429b3] [Medline: 19996978]
29. Gwadz M, Cleland CM, Applegate E, Belkin M, Gandhi M, Salomon N, Heart to Heart Collaborative Research Team. Behavioral intervention improves treatment outcomes among HIV-infected individuals who have delayed, declined, or discontinued antiretroviral therapy: a randomized controlled trial of a novel intervention. AIDS Behav 2015 Oct;19(10):1801-1817. [doi: 10.1007/s10461-015-1054-6] [Medline: 25835462]
30. Gardner EM, McLees MP, Steiner JF, Del RC, Burman WJ. The spectrum of engagement in HIV care and its relevance to test-and-treat strategies for prevention of HIV infection. Clin Infect Dis 2011 Mar 15;52(6):793-800 [FREE Full text] [doi: 10.1093/cid/ciq243] [Medline: 21367734]
31. Chapman DJ, Damio G, Young S, Pérez-Escamilla R. Effectiveness of breastfeeding peer counseling in a low-income, predominantly Latina population: a randomized controlled trial. Arch Pediatr Adolesc Med 2004 Sep;158(9):897-902. [doi: 10.1001/archpedi.158.9.897] [Medline: 15351756]
32. Kelly JA, St Lawrence JS, Diaz YE, Stevenson LY, Hauth AC, Brasfield TL, et al. HIV risk behavior reduction following intervention with key opinion leaders of population: an experimental analysis. Am J Public Health 1991 Feb;81(2):168-171. [Medline: 1990853]


## Abbreviations

**AIDS:** acquired immune deficiency syndrome
**ART:** antiretroviral therapy
**CARE+:** Computer Assessment & Rx Education for HIV-positives
**HIV:** human immunodeficiency virus
**ICT:** information and communication technology
**PLHA:** people living with HIV/AIDS
**RCT:** randomized controlled trial
**TAM:** Technology Acceptance Model
**VAS:** visual analog scale



XSL•FO
RenderX

*Edited by G Eysenbach; submitted 04.04.16; peer-reviewed by N Abdala, K Claborn; comments to author 08.06.16; revised version received 22.06.16; accepted 23.06.16; published 13.07.16.*

*Please cite as:*
*Kurth AE, Chhun N, Cleland CM, Crespo-Fierro M, Parés-Avila JA, Lizcano JA, Norman RG, Shedlin MG, Johnston BE, Sharp VL Linguistic and Cultural Adaptation of a Computer-Based Counseling Program (CARE+ Spanish) to Support HIV Treatment Adherence and Risk Reduction for People Living With HIV/AIDS: A Randomized Controlled Trial*
*J Med Internet Res 2016;18(7):e195*
*URL: http://www.jmir.org/2016/7/e195/*
*doi:10.2196/jmir.5830*
*PMID:27417531*

©Ann E Kurth, Nok Chhun, Charles M Cleland, Michele Crespo-Fierro, José A Parés-Avila, John A Lizcano, Robert G Norman, Michele G Shedlin, Barbara E Johnston, Victoria L Sharp. Originally published in the Journal of Medical Internet Research (http://www.jmir.org), 13.07.2016. This is an open-access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/2.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on http://www.jmir.org/, as well as this copyright and license information must be included.



XSL•FO
RenderX

Original Paper

# Being an Informed Consumer of Health Information and Assessment of Electronic Health Literacy in a National Sample of Internet Users: Validity and Reliability of the e-HLS Instrument

Gül Seçkin[1], PhD; Dale Yeatts[1], PhD; Susan Hughes[1], MA; Cassie Hudson[1], MA; Valarie Bell[2], MA

[1]University of North Texas, Denton, TX, United States

[2]Texas Woman's University, Denton, TX, United States

**Corresponding Author:**
Gül Seçkin, PhD
University of North Texas
1155 Union Circle #311157
Denton, TX, 76203
United States
Phone: 1 (940) 565 2296
Fax: 1 (940) 369 7035
Email: gul.seckin@unt.edu

## Abstract

**Background:** The Internet, with its capacity to provide information that transcends time and space barriers, continues to transform how people find and apply information to their own lives. With the current explosion in electronic sources of health information, including thousands of websites and hundreds of mobile phone health apps, electronic health literacy is gaining an increasing prominence in health and medical research. An important dimension of electronic health literacy is the ability to appraise the quality of information that will facilitate everyday health care decisions. Health information seekers explore their care options by gathering information from health websites, blogs, Web-based forums, social networking websites, and advertisements, despite the fact that information quality on the Internet varies greatly. Nonetheless, research has lagged behind in establishing multidimensional instruments, in part due to the evolving construct of health literacy itself.

**Objective:** The purpose of this study was to examine psychometric properties of a new electronic health literacy (ehealth literacy) measure in a national sample of Internet users with specific attention to older users. Our paper is motivated by the fact that ehealth literacy is an underinvestigated area of inquiry.

**Methods:** Our sample was drawn from a panel of more than 55,000 participants maintained by Knowledge Networks, the largest national probability-based research panel for Web-based surveys. We examined the factor structure of a 19-item electronic Health Literacy Scale (e-HLS) through exploratory factor analysis (EFA) and confirmatory factor analysis, internal consistency reliability, and construct validity on sample of adults (n=710) and a subsample of older adults (n=194). The AMOS graphics program 21.0 was used to construct a measurement model, linking latent factors obtained from EFA with 19 indicators to determine whether this factor structure achieved a good fit with our entire sample and the subsample (age ≥ 60 years). Linear regression analyses were performed in separate models to examine: (1) the construct validity of the e-HLS and (2) its association with respondents' demographic characteristics and health variables.

**Results:** The EFA produced a 3-factor solution: communication (2 items), trust (4 items), and action (13 items). The 3-factor structure of the e-HLS was found to be invariant for the subsample. Fit indices obtained were as follows: full sample: $\chi^2$ (710)=698.547, *df*=131, *P*<.001, comparative fit index (CFI)=0.94, normed fit index (NFI)=0.92, root mean squared error of approximation (RMSEA)=0.08; and for the older subsample (age ≥ 60 years): $\chi^2$ (194)=275.744, *df*=131, *P*<.001, CFI=0.95, NFI=0.90, RMSEA=0.08.

**Conclusions:** The analyses supported the e-HLS validity and internal reliability for the full sample and subsample. The overwhelming majority of our respondents reported a great deal of confidence in their ability to appraise the quality of information obtained from the Internet, yet less than half reported performing quality checks contained on the e-HLS.

***(J Med Internet Res 2016;18(7):e161)*** doi:10.2196/jmir.5496



XSL•FO
RenderX

**KEYWORDS**

health literacy; health information technology; Internet; information; ehealth

## Introduction

Technological advancements inevitably change the information dissemination process by creating new information outlets and developing a platform for new sources [1]. The emergence of Web 2.0 has changed the way consumers interact with technology, information, and health providers. Electronic health, health-related Internet-based technology, and information and communication technologies are broad terms encompassing an array of electronic and mobile phone apps that uses the Internet to deliver health and medical information, independent of space and time considerations often associated with more conventional sources of information. People use desktops, laptops, tablets, and smart phones to access information. These technologies are closely interwoven with the medical field altering self-health care behavior by transforming the scope, breadth, and pace with which information is obtained [2,3]. A study performed in the United States in 2012 found that 81% of the Internet users searched the Web for health information [4], with the majority looking for information about a specific condition or disease [5]. According to the 2014 Pew Internet survey, approximately 1 in 4 people with a chronic illness have read someone's posting about a health issue on a website. Over 70% of people in Europe access health information on the Internet [6]. Studies have also reported that the Internet-based information has a strong effect on how people manage their health. Specifically, Americans often turn to health information on the Internet before seeing a health professional [7]. In fact, people now use the Internet more often than consulting with their doctors [8]. Underlying the growing use of the Internet to gather information is a willingness to become involved in health care decision making and the ability to make informed choices and decisions [9]. As Dutta-Bergman stated, "the critical role of the Internet as a health information resource has shifted traditional patterns of consumer health information use, the physician-patient relationship, and health services delivery" [10]. Numerous scholars have discussed the transformative effect of the Internet on our self-care transforming patients into a reflexive consumer who can make informed decisions. Ehealth information resources have empowered patients to make informed decisions by improving their ability to communicate with their health care providers [11-13].

The term ''e-patients'' was coined to describe individuals who are empowered by various technology-based health information tools and apps, but concerns persist about information accuracy, credibility, and quality [14]. Considerable health information available on the Internet is of varying quality; much of it may be oversimplified, incomplete, inaccurate, or misleading [15]. Although it has been shown that patient–physician interactions can prove more satisfactory thanks in part to better informed patients, nearly 60% of ehealth seekers report that they have hesitated talking to their providers about information from the Internet due to fears of straining their relationship with their physician [16]. Moreover, most people fail to apply any criteria to assess the quality of Web-based information, and instead, they trust that source is credible [17]. Complicating the issue is the fact that according to the Institute of Medicine, nearly 90 million Americans have low health literacy, adversely affecting their ability to appraise health information before making and implementing health care decisions [18]. Most Internet health information searches are generally conducted through a general search engine, accessing a multitude of websites of varying quality. Not surprisingly, as Web-based sources of information proliferate, people report increasing confusion and uncertainty about the quality of information available [19]. Fortunately, there are a number of general guidelines for appraising the credibility and quality of ehealth information, including measures of content accuracy, the provision of disclosure statements, and the currency of information, which constitute ehealth literacy skills. Unfortunately, most users are more influenced by the design and appeal of a website when determining its trustworthiness [12]. Consequently, this raises the concern that ehealth seekers might engage in behavioral practices that might be harmful or dangerous to their health. Information overload, sifting through vast amounts of information while simultaneously trying to decipher its quality has been described as mindboggling and may lead to negative affect, such as fear or anxiety.

Although Internet use may lead to a sense of patient empowerment, empowerment without the requisite high level of health literacy may pose a health risk should a patient misuse the information or decide there is no real need to see a doctor [20]. Thus, the American Medical Association and the National Committee on Quality Assurance have recommended ehealth literacy as one of the top areas for national action. The emphasis on the importance of health promotion and patient self-care in maintaining health and well-being, and a partnership with providers via access to information technology, has led to increased professional discourse on the value of ehealth literacy [18-22]. As the role of digital information technologies in health research continues to unfold, it is necessary to examine the synergy between the multidimensional factors associated with health literacy and their effects on self-health care outcomes [23].

The Institute of Medicine considers health literacy to represent a "constellation of skills" necessary to act on health care information [18]. The lack of an integrated theoretical framework has led researchers to operationalize health literacy in different ways, leading to limited progress in understanding and measuring health literacy [24,25]. Traditionally, health literacy was defined as an individual's capacity to obtain, process, and understand basic health information and the services needed to make appropriate health decisions [18]. The US National Assessment of Adult Literacy defined health literacy as ''the ability of U.S. adults to read, understand, and apply health-related information presented in written English to function in society and achieve one's goals.''[26]. Ratzan et al, define health literacy as "[t]he degree to which individuals can obtain, process, understand, and communicate about health-related information needed to make informed health



XSL•FO
RenderX

decisions" [3]. Sørensen's definition encompasses the individual's competence to appraise health information to make judgments concerning personal health care [26]. These definitions provided a foundation for a concept of health literacy that evaluated reading and medication-related numeracy skills. Several other researchers' definition includes communication with health care providers [27-28]. Thus, health literacy is associated with self-care, health promotion and wellness, and better navigation of the health care system, all of which are important for regaining and maintaining health [29-30]. Health literacy is particularly salient to the aging population's increasing longevity. Older adults develop substantial health care needs as they age, and research examining this demographic group's health literacy is much needed. Moreover, the burgeoning focus on successful aging strategies and quality of life for the ageing has provided a greater impetus for understanding their capacity to make informed appraisals of Internet health information [31,32]. Notwithstanding, despite the continual evolution of health literacy, there has been a research lag in establishing multidimensional instruments to measure the construct [33]. In our research, we emphasize the importance of measuring ehealth literacy because most existing measures fail to target the health literacy skills of ehealth information seekers despite its potential impact on a range of health outcomes [9,33]. As Pleasant stated "many people have offered many definitions, yet those definitions have not been formally or fully tested. While health literacy evaluative and measurement tools often claim to be based on one definition or another, the specific constructs within the definitions have rarely, if ever, been explicitly built into and tested with the evaluative tool.'' [34]. Furthermore, the definition of health literacy needs to be broadened [35]. In the age of electronic health information, measures that reflect an ability to read static text still predominate. Norman emphasized the need for research to develop a valid self-report measure that assesses ehealth literacy skills due to the expansion of health information available on the Internet [36]. There are few tools to assess users' ability to engage in ehealth in an informed way [37-38]. ''Lacking empirical evidence of the relationship between different literacy skills, reading and numeracy skills are often used as proxies of literacy in research and practice'' [39]. However, the ehealth websites require a consumer to have the ability to appraise the content of informational resources to make health decisions. "Accessing health information had never been easier than in the current information age as the Internet's vast content and global reach allows health consumers to quickly connect with the latest information'' [40]. Although health literacy for written resources is well defined and various measures exist, the evaluation of health literacy in a digital context is less clear. Obtaining health information from the Internet can be very helpful to those who are able to discern the difference between reliable and unreliable health information websites. Accordingly, a new generation of health literacy assessment tools, as a response to the new digital technologies, is needed. The measurement of health literacy should be adjusted to reflect technological changes [41].

Our goals in this research were twofold: (1) to develop a tool to be used in ehealth literacy research and to examine its psychometric properties; and (2) to help understand how ehealth literacy is associated with health care variables. Our ehealth literacy measure is a tool designed to assess the degree to which people possess the skills required to use ehealth information in an informed way. Originally, Norman and Skinner introduced the concept of ehealth literacy, defining it as the ability to seek, find, understand, and appraise health information from digital sources and apply this knowledge to solve health problems. Their Lily Model included 6 core health literacy skills depicted as petals of a lily: traditional (reading ability and numeracy), information, media, health, computer, and science [41]. According to Jordan et al, most of previous instruments assessed user competency with Web technologies; however, they failed to capture user skills required in the age of ehealth information through the Internet. They have also been found to have substantial psychometric weaknesses [42]. These measures, such as the Rapid Estimate of Adult Literacy in Medicine and the Test of Functional Health Literacy in Adults among others, assess operational skills (basic skills needed to use the Internet), formal skills navigation, information skills (locating information), logic skills (ability to understand information), functional literacy (reading and understanding health information), and strategic skills (applying information to health problems) [43-49]. The 8-item eHealth Literacy Scale designed by Norman and Skinner measures a consumer's perceived skills at using information technology such as their comfort in using computers and ability to locate health information [41]. Although these instruments address a combination of technical aspects related to the use of the Internet and content provided, they do not measure ability to appraise health information. A research study reported that almost 90% of the participants in a discussion of health literacy agreed that current measures of health literacy do not match with the current understanding of health literacy in age of information technology [24]. Understanding ehealth literacy requires an examination of critical issues such as the users' ability to find appropriate information and use it to gain better control over their personal health. Although current assessments of health literacy focus primarily on reading ability, our review of the literature suggested the need for updated measures of health literacy that would measure information search strategies and skills to judge the quality of information found [12,50,51]. Current research instruments fail to capture important aspects of ehealth literacy such as appraisal, trust, and the communicative aspects of it as an interactive process. To address this gap, we designed items to reflect these components of ehealth literacy.

This study advances this effort by developing an ehealth literacy scale for users of digitally provided health information. Most existing measures of health literacy focus on a single dimension, which tends to be a reading comprehension test emphasizing a relatively narrower cognitive capacity to understand health-related texts and materials [24]. The need to navigate health websites with confidence is particularly important because the consequences for using low-quality, misleading, or false information could endanger health and possibly result in death [35]. Through our review, we identified key attributes of ehealth literacy demands. An area of consensus is evaluating information to discern high-quality information from low-quality information. Accordingly, our measure and its items reflect this area of consensus.



XSL•FO
RenderX

## Methods

### Data Source and Sample

The sample consists of respondents who used the Internet for health information (N=710). The Knowledge Networks (KN), a nonprofit academic research firm, recruited the respondents who are members of the first Web-based panel representative of the US population. The KN Panel consists of about 50,000 US residents, aged 18 years or older. The KN uses an address-based sample frame derived from the US Postal Service Delivery Sequence File, which covers 97% of US households, thereby maximizing sample representativeness. Address-based sampling permits probability-based sampling of addresses including those households that have unlisted telephone numbers, do not have landline telephones, do not have Internet access, and do not have devices to access the Internet. Respondents are randomly selected, in contrast to the opt-in convenience sampling design of most other Web-based panels. The KN Panel members who were randomly selected were invited to become panel members. For those selected households that do not have Internet access or devices to access the Internet, we provided a Web-enabled computer with free Internet service to enable their participation as Web-based panel members. The KN obtained the participants' consent before they become panel members [52].

For this study, 1315 participants were randomly selected after being contacted via an email. Potential participants were prescreened through the question "Do you seek health or medical information on the Internet for yourself and for others?" We obtained a 70% response rate and received a total of 710 completed Web-based questionnaires. The Web-based survey consisted of 50 questions. It was self-administered and accessible for a designated period of time. Respondents were able to complete the survey only once. The 19-item electronic health literacy measure was developed through an extensive multistep scale development and evaluation process. We created items based on a review of the literature. Approval of the Institutional Review Board of the University of Maryland, Baltimore County, was also obtained (protocol number: Y11GS21145) before the study's launch. During the pilot phase of our project, we tested general readability and item wording. We field-tested the items (n=10) to assess clarity of wording and general readability of the items and whether participants interpreted the items as we intended. No problems were reported by our pilot study respondents in regard to clarity of survey questions.

An inherent part of any survey is nonresponse. The KN attains a 65% to 70% survey completion rate as opposed to 2% to 16% for opt-in Web-based panels. Our specific survey sample was drawn at random from the panel members who were randomly recruited in accordance with scientifically accepted sampling theory and methods. Accordingly, our specific survey sample represents a simple random sample from the larger probability-based panel designed to be statistically representative of the US population. Because all KN Panel households were selected randomly with a known probability of selection and because our survey-specific panelists were then also randomly selected from the larger panel, our results can be interpreted with the statistical confidence relative to the population of the United States [52].

Furthermore, the KN states,

> *in certain cases, a survey sample calls for pre-screening, that is, members are drawn from a subsample of the panel. There are also several sources of survey error that are an inherent part of any survey process, such as non-coverage and non-response due to panel recruitment methods and to inevitable panel attrition. We address these sources of sampling and non-sampling error by using a panel demographic post-stratification weight as an additional adjustment based on demographic distributions from the most recent data from the Current Population Survey (CPS). This weighting adjustment is applied prior to the selection of any client sample from KnowledgePanel, and these weights are used in the stratified, weighted, selection procedure for drawing samples from the panel. All the above weighting is done before the study sample is drawn. Once a study sample is finalized, a set of study-specific post-stratification weights are constructed so that the study data can be adjusted for the study's sample design and for survey nonresponse. Starting with each panel member's base weight, an iterative raking procedure is used to achieve an optimal approximation of the relevant benchmarks to make survey respondents representative [52].*

### Statistical Analysis

Our psychometric analyses started with exploratory factor analysis (EFA) using principal component analysis and varimax rotation to identify these theorized latent dimensions represented in the variables and to define the underlying structure among the variables. This enabled us to have an initial confidence in our conceptualization. As Hair et al [53] wrote "[e]xploratory factor analysis can be performed to provide a preliminary check on the number of factors and the pattern of loadings. Then proceed to a confirmatory test of measurement theory (to establish the construct validity of the newly designated scale)." We examined how many factors existed, whether factors were correlated, and which variables best measured each factor. This also enabled us to determine whether any underlying structure existed for measures on the 19 variables. Hair et al [53] wrote, CFA cannot be conducted appropriately unless the researcher can specify both the number of constructs that exist within the data to be analyzed and which specific measures should be assigned to each of these constructs. After performing EFA, we proceeded with confirmatory factor analysis (CFA) to determine whether the items in our instrument support the 3-factor structure, which provided evidence that the item measures taken from our sample represent the true score that exists in the population. Beginning analytical procedures with EFA by examining the measurement model followed by CFA was also reported in the literature on the psychometric validation of new instruments [54-57]. The factor structure proposed by EFA on the full sample was validated with a subsample comparison



XSL•FO
RenderX

approach using a sample of older adults (age ≥ 60 years). This enabled us to assess the stability of the factor structure of the electronic Health Literacy Scale (e-HLS).

We, then, proceeded with CFA for the full sample and subsample. First, the AMOS graphics program was used to construct an input path diagram representing the measurement model that linked the ehealth literacy factors (latent variables) with e-HLS indicators. The model included covariances between the 3 factors, previously proposed by EFA. Data were entered for 710 cases, standardized beta coefficients were generated for all regressions of indicator variables on factors that were included in the model, and the covariance between the factors were obtained. The $R^2$ values for all 19 e-HLS indicators were also generated. We repeated these analytical procedures for our subsample by entering data for 194 cases separately from the full sample. The chi-square significance test and overall model fit indices were estimated including the comparative fit index (CFI), the normed fit index (NFI), and root mean squared error of approximation (RMSEA).

Item total correlations and Cronbach alpha internal consistency reliability coefficients were calculated for the full sample and subsample. We tested the validity of our scale by examining its correlations with respondents' demographic characteristics: age, gender, race or ethnicity, marital status, education level, and income. We performed ordinary least squares regression analyses to examine our scale's construct validity. We regressed the composite scale on variables in our dataset that are conceptually and empirically related to health literacy. These variables were as follows: perceived empowerment, health interactions, health communication, experiencing health problems, noncompliance, and negative effect. All linear regression analyses were controlled for demographic covariates. Mean replacement procedure was used when missing data are less than 2% of responses for an item. We used SPSS 21.0 in our analyses.

## Measures

Our measure, its dimensions, and the items representing the dimensions were constructed from a literature review of health literacy materials in the Medline, PsycInfo, ERIC, Sociological Abstracts, and Web of Science databases. We also reviewed existing instruments developed for print and Web-based health information materials. We conducted a comprehensive literature review to identify key skills associated with health literacy. In this literature review, we examined how literacy demands of digital health information materials are related to evaluation of information quality. Our review revealed that most existing tools target traditional health literacy for print resources. Given this constraint, we decided to create items based on our review of the literature. We generated items to operationalize each of the 3 conceptual domains identified in the literature: trust, action and behavior. Because the concept of ehealth literacy is increasingly conceptualized as consisting of skills related to evaluating, communicating, and using that information to make informed decisions, we designed our item to reflect these skills.

To measure participants' trust in the Internet-based sources of health information, actions they take to evaluate information, and the extent to which they engage in informational exchange with health professionals, we asked them to indicate their agreement with the items of our measure. The theoretical basis for the trust items is literature on trustworthiness of Web-based health information such as the California Health Care Foundation's report and other related literature [58-62]. Scale items designed to measure the communication dimension are based on the findings of previous studies of patient–provider dialogue [63-65]. The items we theorized to represent the action dimension are derived from a review of literature on uses of the Internet for health information and how Internet users evaluate information including the Medical Library Association's guidelines and other related publications [66-70]. The following is a list of specific items we used in our research:

*Demographic and socioeconomic covariates* included age, race or ethnicity, gender, marital status, education level, and income. Age was measured as an ordinal variable. Gender was coded as (0) male and (1) female. Response categories for race or ethnicity and marital status were collapsed to account for small cell sizes and were measured as dichotomous variables. Race or ethnicity was measured as (0) Caucasian and (1) minority. Education level was coded as (1) high school or less, (2) some college or associate degree, (3) college degree, and (4) postgraduate degree. Annual family income was categorized into 4 groups: (1) $29,999 or less, (2) $30,000 to $59,999, (3) $60,000 to $99,999, and (4) $100,000 and above. Marital status was measured as (0) married and (1) nonmarried.

*Electronic health literacy* was measured with our scale, which we labeled e-HLS. It is a 19-item self-report scale that examines the (1) behavioral, (2) communicational, and (3) attitudinal components of health literacy among ehealth information seekers. Each item was rated on a 5-point Likert scale ranging from 1="never or strongly disagree" to 5="always or strongly agree." The survey assessed whether ehealth information seekers do the following when gathering information from the Internet: (1) read disclosure statements on health websites; (2) check for credentials and institutional affiliations of those who provide information on websites; (3) check the ownership of a health website; (4) check a website's sponsor(s); (5) check for financial ties between website information and the website's sponsor(s); (6) appraise the adequacy and integrity of information providers' credentials; (7) check to see whether a physical address is provided; (8) check for stated goals and objectives; (9) appraise whether coverage of health topics is clear and comprehensive; (10) check whether other print or Web resources confirm information provided; (11) checked whether information is current and updated; (12) check for the last time information was updated. We also asked (13) if they were confident in their ability to appraise information quality on the Internet; and if they (14) asked health professionals for advice about where to find credible information on the Internet; (15) discussed information obtained from the Internet with a health professional; (16) believed information provided on the Internet was credible; (17) believed information provided on the Internet was balanced and accurate; (18) thought information provided on the Internet was the same as or better than what most health professionals provided; and (19) trusted the Internet for obtaining accurate health information. We reverse-coded the



XSL•FO
RenderX

last 4 items so that lower scores represent greater consistency with awareness of varying quality of health information. Our scale had a Cronbach alpha coefficient of .93.

All our questionnaire items had equal weight and were measured on the same metric, a 5-point Likert measurement scale. This ensured that none of the items were more influential than the other items in averaging an overall score for our scale. Consistent with the literature, we calculated a score for each subscale that used items with different response options and performed separate reliability and validity analyses for each. DeCoster stated, "[y]ou might create a group of items to determine respondents' opinions on each of these issues. Sometimes a single questionnaire contains items from several different scales mixed together. This is perfectly legitimate. In this case your items making up different subscales will be slightly different'' [71]. In fact, it is not unusual for instruments with subscales to include items with different response options. The expectation is that the direction of the magnitude of the responses between items should be consistent throughout the scale. In other words, questions should be written to indicate that higher scores should indicate more positive responses or greater magnitude on the variable and vice versa [56,71].

*Positive health interaction* was measured by asking respondents to indicate the extent to which they agreed with the following statements: (1) "I receive more attention to my questions from health providers as a result of gathering information from the Internet," (2) "I receive more information from health providers as a result of gathering information from the Internet," and (3) "Interactions with health providers have become more respectful as a result of gathering information from the Internet." Response options ranged from 1="strongly disagree" to 5="strongly agree." The Cronbach alpha reliability coefficient is .87.

*Strained Health Interaction* was measured by asking respondents to indicate the extent to which they agreed with the following statement: "Interactions with health providers have become strained as a result of bringing in health and medical information from the Internet to my appointments." Response options ranged from 1="strongly disagree" to 5="strongly agree." We first reverse-coded the item and included it with the rest of positive health interaction items, but, we found the Cronbach alpha reliability value to be less than the threshold value of .70. Alpha if item deleted analysis suggested dropping this strain item. Thus, we separated it from the rest of health interaction items and performed a single-item analysis.

*Health communication* was measured through questions that asked respondents to indicate the extent to which they agree with the following statements: (1) "Information on the Internet helps me to communicate more effectively with health providers during appointments," (2) "Information on the Internet helps me to ask more informed questions to health providers," and (3) "Information on the Internet helps me to better understand what my health provider is telling me during appointments." Response options ranged from 1="strongly disagree" to 5="strongly agree." The Cronbach alpha reliability is .88.

*Nonadherence* was assessed through the following questions: (1) "Do you change your willingness to accept a health care provider's treatment after reading information on the Internet?," (2) "Do you doubt diagnosis or treatment of a health care provider if it conflicts with information on the Internet?,'' and (3) "Have you ever changed a health care provider's treatment as a result of information obtained from the Internet?" Response options ranged from 1="never" to 5="always." The Cronbach alpha reliability coefficient for this measure is .71.

*Perceived empowerment* was assessed with a single item that asked respondents to indicate the extent to which they agreed with the following statement: "Gathering information from the Internet about my health makes me feel empowered." The response options ranged from 1="strongly disagree" to 5="strongly agree."

*Negative effect* was measured with a statement that asked respondents to indicate the extent to which they agreed with the following statement: "Gathering information from the Internet about my health makes me worried and/or anxious." The response options ranged from 1="strongly disagree" to 5="strongly agree."

*Health problem* was measured with the following question: "Have you ever experienced a health problem as a result of using the Internet information?" Response options ranged from 1="never" to 5="always."

## Results

Our sample consisted of adults (n=710), almost equally distributed between men and women (381/710, 53.7% women), between the ages of 18 and 93 years with a mean of 48.82 ± 16.43. About 68% (481/710) were married, and 543 of 710 (77%) were Caucasian. Almost 40% (265/710) had a college degree or higher, and 405 of 710 (57%) earned $60,000 or more. Our comparison subsample consisted of respondents who were aged 60 years or older. They made up almost 30% (n=194) of our sample. About 40% (73/194, 37.6%) of them had a college degree or higher, and slightly more than half (99/194, 51.1%) reported an income level of $ 60,000 or more. Just over 60% were married (121/194, 62.4%), and a little over 80% (160/194, 82.5%) were Caucasian. We examined whether respondents' sociodemographic characteristics and ehealth literacy were associated. There was a significant mean difference for the communication factor between men and women (2.10 vs 2.24, $P$=.047). There was also a racial or ethnic difference in means reported for the action factor with Caucasian participants reporting higher scores on the e-HLS (2.52 vs 2.35, $P$=.05). Significantly higher means were reported for the action and communication factors by those with higher education (2.20 vs 2.83, $P$=.001 for the action factor and 1.96 vs 2.34, $P$==.001 for the communication factor). There is also a significant mean difference on the overall e-HLS score between respondents who had higher levels of education compared with those with lower levels (2.13 vs 2.77, $P$=.001). However, there was no statistically significant difference in the trust factor based on education (2.84 vs 2.71, $P$=.32). Respondents with higher income levels were also found to have a higher score on the overall e-HLS than those with lower incomes (2.52 vs 2.14, $P$=.01). We found a significant association between ehealth literacy and respondent age at neither the item nor factor level. Finally, married respondents had higher averages for the communication factor



XSL•FO
RenderX

than nonmarried respondents (2.22 vs 2.08, *P*=.05). We examined bivariate associations of ehealth literacy with health-related variables in our survey. Respondents with higher scores on our measure of electronic health literacy reported higher sense of perceived empowerment (r=.395, *P*=.001), lower negative effect (worry or anxiety; r=-.116, *P*=.002), perceptions of more positive health care interactions with providers (r=.290, *P*=.001), and better health care communication (r=.427, *P*=.001). However, we also found a significant positive association with nonadherence (r=.454, *P*=.001) and experiencing a health problem as a result of using the Internet-based information (r=.128, *P*=.001). No significant associations were found between the ehealth literacy and perceived strain in health care interactions.

Next, we performed univariate examination of our scale items (Table 1). The item means for the full sample ranged from 1.93 ± 1.08 to 3.24 ± 0.95. The highest mean score was 3.24 ± 0.95 for the item that inquired about perceived confidence in ability to appraise information quality, followed by the item that asked about the extent to which the respondents believed that the Internet provides high-quality and credible health information (3.09 ± 0.75). The lowest average was found for the item that inquired whether respondents asked health professionals for advice about where to find credible information on the Internet (1.93 ± 1.08). Similar results were obtained for our subsample of older respondents. The highest mean scores were 3.23 ± 0.93 and 3.08 ± 0.74 for the aforementioned items. Similarly, the lowest average was for the item that questioned whether respondents asked health professionals for advice on where to find credible health information on the Internet (1.80 ± 0.98). The mean score of the composite scale was 2.51 ± 0.77 for the full sample, and it was 2.53 ± 0.81 for our subsample. Frequencies and percentages in Table 1 reflect the answers of those who responded “sometimes” ‘most of the time’ and “always.” Standard deviations are shown in parentheses.

Next, we performed the Kaiser–Meyer–Olkin (KMO) measure of sampling adequacy and Bartlett Test of Sphericity (BTS) to determine whether our data were suitable for EFA. Kaiser–Meyer–Olkin value of above 0.50 is needed before proceeding with EFA, whereas values of 0.80 or above are considered very good. A statistically significant BTS (*P*=.05 *)* indicates that sufficient correlations exist among the variables to proceed. The KMO and BTS results in our research indicated that the dataset satisfied the psychometric criteria for EFA analysis. Kaiser–Meyer–Olkin analysis yielded an index of 0.93, and BTS yielded 838.82, *P*=.001. We performed EFA with principal component analysis and varimax rotation using the following criteria: (1) eigenvalue greater than 1, (2) items loading on the same factor (≥0.30), (3) no crossloading, (4) Cattell’s scree test, and (5) conceptual interpretability of factors. Principal component analysis was chosen as a data extraction method because as Hair et al stated “This method focuses on extracting the minimum number of factors to account for the maximum portion of the total variance represented in the original set of variables’’ in the dataset [53]. The varimax rotation converged in 5 iterations. Three factors with eigenvalues greater than 1 emerged from the analyses. The eigenvalues for these factors were as follows: 8.52, 2.74, and 1.03. All the 3 factors explained 65% of the variance. A factor solution that accounts for 60% of the total variance is considered satisfactory [53].

Reestimation of the factor structure in our subsample confirmed this 3-factor solution. The varimax rotation converged in 4 iterations and provided the following eigenvalues: 9.17, 2.67, and 1.05. These factors explained 65% of variance in the data for the subsample of older adults. Although we used no crossloading as one of the criteria in determining the underlying factor structure of the e-HLS, we found that 2 items (perceived confidence to appraise information and discussing information with a health professional) crossloaded with 2 factors. We considered several alternative solutions to ensure that we had identified the best structure (1 less and 1 more factor than the initial solution suggested by EFA). We, then, determined to keep these 2 items in our composite scale because we deemed that they are conceptually important components of ehealth literacy. This decision is based on statisticians’ recommendation that “it is left up to the researcher to be the final arbitrator as to the form and appropriateness of a factor solution, and such decisions are best guided by conceptual rather than empirical bases” [53]. The distributions of the survey items to the factors in our full and subsample are summarized in Table 2 and Table 3, respectively.



XSL•FO
RenderX

**Table 1.** Univariate description of the e-HLS items.

| Scale items | Means and standard deviations | | Item frequencies and percentages | |
|---|---|---|---|---|
| | Full sample | Subsample | Full sample | Subsample |
| Action factor | 2.48 (0.99) | 2.52 (1.05) | | |
| Read disclosure statements | 2.32 (1.20) | 2.54 (1.20) | 44.2% (311) | 54.4% (105) |
| Check credentials and affiliations of author | 2.48 (1.36) | 2.49 (1.42) | 46.3% (327) | 47.7% (92) |
| Check who owns the website | 2.41 (1.40) | 2.22 (1.42) | 43.8% (309) | 34.2% (66) |
| Check who sponsors the website | 2.40 (1.37) | 2.36 (1.42) | 44.3% (311) | 40.3% (76) |
| Check if there is a financial tie between information and sponsor | 2.20 (1.36) | 2.25 (1.41) | 36.4% (257) | 36.3% (70) |
| Appraise whether information provider's credentials seem adequate | 2.54 (1.40) | 2.66 (1.42) | 49.8% (348) | 51.6% (99) |
| Check whether an address is listed on the website | 1.96 (1.09) | 2.08 (1.21) | 29.1% (205) | 31.8% (61) |
| Check whether goals and objectives of the website are clearly stated | 2.26 (1.20) | 2.34 (1.27) | 41.4% (292) | 42.9% (82) |
| Appraise whether there is a clear and comprehensive coverage of the topic | 2.63 (1.33) | 2.75 (1.36) | 54.2% (379) | 55.6% (105) |
| Check whether other print or Web resources confirm the information | 2.57 (1.33) | 2.57 (1.26) | 51.5% (363) | 51.3% (99) |
| Check whether information is current and updated recently | 2.90 (1.35) | 2.91 (1.40) | 61.5% (432) | 60.6% (117) |
| Check whether the last update of information is prominent on the website | 2.66 (1.33) | 2.68 (1.38) | 53.4% (374) | 51.8% (98) |
| Confident of being able to appraise information quality on the Internet | 3.24 (0.95) | 3.23 (0.93) | 80.8% (566) | 77.6% (149) |
| Trust factor | 2.79 (0.64) | 2.78 (0.65) | | |
| Trust the Internet to provide accurate information | 2.72 (0.86) | 2.74 (0.87) | 61.5% (432) | 62.8% (120) |
| Think information on the Internet as credible | 3.09 (0.75) | 3.08 (0.74) | 84.5% (592) | 85.4% (164) |
| Think information on the Internet as balanced and accurate | 2.95 (0.73) | 2.94 (0.74) | 79.2% (557) | 78.1% (150) |
| Think information on the Internet better than what most health providers supply | 2.41 (0.87) | 2.35 (0.87) | 46.9% (328) | 41.3% (79) |
| Communication factor | 2.18 (0.90) | 2.10 (0.87) | | |
| Discuss the information with a health provider | 2.42 (1.07) | 2.40 (1.06) | 49.9% (311) | 46.9% (90) |
| Ask a health provider where to find credible information on the Internet | 1.93 (1.08) | 1.80 (0.98) | 29.2% (206) | 24.3% (47) |



XSL•FO
RenderX

**Table 2.** Factor analysis of the full-sample e-HLS items.

| Scale items | Full sample (n=710) | | | |
|---|---|---|---|---|
| | Factor I | Factor II | Factor III | Item total correlations |
| Read disclosure statements | 0.66 | 0.01 | 0.26 | 0.60 |
| Check credentials and affiliations of author | 0.79 | 0.14 | 0.16 | 0.72 |
| Check who owns the website | 0.79 | 0.12 | 0.23 | 0.74 |
| Check who sponsors the website | 0.84 | 0.16 | 0.14 | 0.78 |
| Check if there is a financial tie between information and sponsor | 0.78 | 0.16 | 0.11 | 0.72 |
| Appraise whether information provider's credentials seem adequate | 0.85 | 0.15 | 0.18 | 0.80 |
| Check whether an address is listed on the website | 0.76 | 0.05 | 0.23 | 0.70 |
| Check whether goals and objectives of the website are clearly stated | 0.79 | 0.03 | 0.13 | 0.74 |
| Appraise whether there is a clear and comprehensive coverage of the topic | 0.83 | 0.03 | 0.11 | 0.77 |
| Check whether other print or web resources confirm the information | 0.80 | 0.12 | 0.08 | 0.77 |
| Check whether information is current and updated recently | 0.85 | 0.08 | 0.05 | 0.77 |
| Check whether the last update of information is prominent on the website | 0.80 | 0.05 | 0.05 | 0.73 |
| Confident of being able to appraise information quality on the Internet | 0.45 | 0.32 | 0.43 | 0.43 |
| Ask a health provider where to find credible information on the Internet | 0.23 | 0.06 | 0.83 | 0.50 |
| Discuss the information with a health provider | 0.35 | 0.19 | 0.57 | 0.54 |
| Trust the Internet to provide accurate information | 0.34 | 0.75 | 0.01 | 0.34 |
| Think information on the Internet as credible | 0.17 | 0.86 | 0.01 | 0.20 |
| Think information on the Internet as balanced and accurate | 0.08 | 0.85 | 0.01 | 0.11 |
| Think information on the Internet better than what most health providers supply | 0.18 | 0.68 | 0.08 | 0.19 |



XSL•FO
RenderX

**Table 3.** Factor analysis of the subsample e-HLS items.

| Scale items | Subsample: (n=194) | | | |
|---|---|---|---|---|
| | Factor I | Factor II | Factor III | Item-Total Correlations |
| Read disclosure statements | 0.64 | 0.13 | 0.33 | 0.72 |
| Check credentials and affiliations of author | 0.81 | 0.02 | 0.12 | 0.73 |
| Check who owns the website | 0.85 | 0.08 | 0.03 | 0.79 |
| Check who sponsors the website | 0.86 | 0.01 | 0.06 | 0.81 |
| Check if there is a financial tie between information and sponsor | 0.82 | 0.04 | 0.06 | 0.76 |
| Appraise whether information provider's credentials seem adequate | 0.89 | 0.01 | 0.07 | 0.84 |
| Check whether an address is listed on the website | 0.74 | 0.04 | 0.29 | 0.93 |
| Check whether goals and objectives of the website are clearly stated | 0.77 | 0.14 | 0.24 | 0.75 |
| Appraise whether there is a clear and comprehensive coverage of the topic | 0.83 | 0.11 | 0.10 | 0.78 |
| Check whether other print or web resources confirm the information | 0.81 | 0.02 | 0.16 | 0.80 |
| Check whether information is current and updated recently | 0.84 | 0.09 | 0.19 | 0.80 |
| Check whether the last update of information is prominent on the website | 0.78 | 0.07 | 0.22 | 0.75 |
| Confident of being able to appraise information quality on the Internet | 0.50 | 0.41 | 0.40 | 0.49 |
| Ask a health provider where to find credible information on the Internet | 0.36 | 0.03 | 0.78 | 0.49 |
| Discuss the information with a health provider | 0.47 | 0.17 | 0.55 | 0.58 |
| Trust the Internet to provide accurate information | 0.16 | 0.81 | 0.13 | 0.39 |
| Think information on the Internet as credible | 0.02 | 0.88 | 0.12 | 0.25 |
| Think information on the Internet as balanced and accurate | 0.09 | 0.86 | 0.08 | 0.09 |
| Think information on the Internet better than what most health providers supply | 0.06 | 0.72 | 0.15 | 0.18 |

On the basis of review of the existing literature, we labeled our first factor as *behavioral literacy (action factor)*. It includes 13 items of behavioral indicators from the e-HLS. The factor loadings ranged from 0.45 to 0.85. The item that inquired if respondents appraised the adequacy of information providers' credentials had the highest factor loading, whereas the item that asked if they were confident of their ability to appraise information quality on the Internet had the lowest factor loading. Similar patterns of factor item loadings emerged in our subsample with factor loadings ranging from 0.50 to 0.89. The factor mean is found to be 2.48 with a standard deviation of 0.99. We identified our second factor as *cognitive literacy (trust factor)*. It consists of 4 items that assessed the perceived accuracy of health information on the Internet. The factor loadings ranged from 0.68 to 0.86. The item that assessed if they believed information provided on the Internet was credible had the strongest factor loading, whereas the item that assessed if they thought information provided on the Internet was the same as or better than what most health professionals provided had the lowest factor loading. Similar factor loading patterns were found for our comparison subsample, with factor loadings ranging from 0.72 to 0.88. The factor mean is 2.79, and standard deviation is 0.64. We identified our third factor as *interactional literacy (communication factor)*. It consists of 2 items that measure the extent to which discussion of Internet information takes place between health care provider and information user. The factor loadings ranged from 0.57 to 0.83. The item that assessed if respondents asked health professionals for advice about where to find credible information on the Internet showed the strongest factor loading, whereas the item that measured if they discussed information obtained from the Internet with a health professional showed the lowest factor loading. A similar pattern emerged with our comparison subsample with factor loadings ranging from 0.55 to 0.78. The mean value for this factor is 2.18 with a standard deviation of 0.91.

Of the 3 factors, the trust factor has the highest mean score (2.79 ± 0.64 for the full sample; 2.78 ± 0.65 for the subsample). The action factor has the next highest mean scores (2.48 ± 0.99; 2.52 ± 1.05). The communication factor has the lowest averages for the full sample and the subsample (2.18 ± 0.90; 2.10 ± 0.87). When the correlations between the factors were examined, low-to-moderate to moderate-to-high significant associations emerged. In the full sample, the action and communication factors have the highest correlation with each other (r=.59, *P*=.001) with the trust factor correlating with both communication and action factors (r=.21, *P*=.001 and r=.17, *P*=.001). Similar results were obtained for our subsample. The action and communication factors have a high correlation with each other (r=.60 at *P*=.001) and the trust factor correlating with both communication and action factors (r=.23, *P*=.001 and r=.17, *P*=.001).

Confirmatory factor analysis using AMOS 21 statistical program verified that the 3-factor structure of the e-HLS is invariant for



XSL•FO
RenderX

the full and subsample and achieved a good fit with both. Comparative fit index and NFI values close to 1 and RMSEA index less than 0.10 are generally deemed to be a good fit. Fit indices obtained for the full sample were as follows: chi-square (710)=698.547, *df*=131, *P*=.001, CFI=0.94, NFI=0.92, RMSEA=0.07. All factor loadings were significant (*P*<.001) with standardized regression coefficients exceeding 0.40. The estimates for the older subsample were as follows: chi-square (194)=275.744, *df*=131, *P*=.001, CFI=0.95, NFI=0.90, RMSEA=0.08. All factor loadings were also significant (*P*<.001) with standardized regression coefficients exceeding 0.50. The significant goodness-of-fit value given by the chi-square index was likely a result of the sample size because as sample size increases, the chi-square value quickly approaches significance and should not be interpreted as an indication of poor model fit [72,73]. The output path diagram showing the computed values for the entire sample is depicted in Figure 1, and that of the subsample is depicted in Figure 2. Finally, internal consistency reliability analysis of the e-HLS demonstrated high Cronbach alpha values: .93 for the full sample and .94 for our subsample.

Next, we examined the validity of our measure by performing external correlates test. According to DeCoster, ‘‘you can (and should) assess validity in a number of different ways. Each time you demonstrate that the scale acts in a way consistent with the underlying construct you make a more convincing argument that the scale provides an accurate representation of that construct’’ [71]. The typical scale validation involves assessing the newly developed scale as it relates to other constructs. Spector stated, “[t]he typical scale-validation strategy involves testing the scale of interest in the context of a set of hypothesized interrelations of the intended construct with other constructs’’ [56]. To confirm the validity of our new scale, we needed to assess how it associated with related constructs. For this purpose, therefore, we performed both bivariate and multivariate analyses using our full sample and subsample.

First, we examined bivariate associations of ehealth literacy with health-related variables in our survey. Respondents with higher scores on our measure of electronic health literacy reported a higher sense of perceived empowerment (r=.395, *P*=.001), less negative effect (worry or anxiety; r=-.116, *P*=.002), perceptions of more positive health care interactions with providers (r=.290, *P*.001), and better health care communication (r=.427, *P*=.001). However, we also found a significant positive association with nonadherence (r=.454, *P*=.001) and experience of a health problem as a result of using Internet-based information (r=.128, *P*=.001). No significant association existed with perceived strain in health care interactions. Our eHealth literacy instrument also revealed that higher scores on our scale were correlated with education (r=.20, *P*=.001) and income (r=.10, *P*=.01). In regard to the subsample, we found that our instrument displayed the following correlations: older respondents who obtained higher scores on our instrument reported a higher sense of perceived empowerment (r=.502, *P*=.001), perceptions of more positive health care interactions with providers (r=.304, *P*=.001), and better health care communication (r=.489, *P*=.001). We also found a significant positive association with nonadherence (r=.533, *P*=.001), strained interactions with health providers (r=-.176, *P*=.01), and less negative effect (r=-.152, *P*=.001). In our subsample, higher scores on our ehealth literacy scale were negatively correlated with older age (r=-.167, *P*=.02) and positively correlated with education (r=.296, *P*=.001).

We also performed linear regression analyses to examine the extent to which the factorial structure of the e-HLS was differentially associated with health variables from our survey dataset. Consistent with the literature cited in the beginning of this paper, we chose the following variables to run our multivariate regression analyses: health interaction, health communication, nonadherence, perceived empowerment, negative effect, and health problem. As summarized in Table 4 and Table 5 (numbers are rounded up), the regression coefficients for the action factor for the full sample were significant for the following variables after controlling for the effect of demographic variables: perceived empowerment (β=.303, *P*=.001), nonadherence (β=.316, *P*=.001), health communication (β=.206, *P*=.001), and negative effect (β=-.174, *P*=.001). For our subsample, we found the action factor to be significantly associated with health communication (β=.140, *P*=.053), perceived empowerment (β=.365, *P*=.001), negative effect (β=-.250, *P*=.005), and nonadherence (β=.312 *P*=.001). The action factor had no significant association with positive health interaction, strained health interaction, and health problem. The *b* represents unstandardized regression coefficients (slope), and β represents standardized regression coefficients. The regression coefficients for the trust factor for the full sample were significant for the following variables: perceived empowerment (β=.293, *P*=.001), nonadherence (β=.216 *P*=.001), positive health interaction (β=.282, *P*=.001), health communication (β=.280, *P*=.001), strained health interaction (β=.092, *P*=.02), and negative effect (β=.077, *P*=.04). Significant associations for the trust factor in the subsample of older adults include positive health interaction (β=.340, *P*=.001), health communication (β=.326, *P*=.001), perceived empowerment (β=.299, *P*=.001), and nonadherence (β=.249, *P*=.001). The trust factor has no significant association with reports of a health problem at the multivariate level. The significant coefficients for the communication factor after controlling for the effects of demographic coefficients include the following variables: perceived empowerment (β=.106, *P*=.001), nonadherence (β=.191, *P*=.001), positive health interaction (β=.308, *P*=.001), health communication (β=.323, *P*=.001), and health problem (β=.147, *P*=.002). In the subsample, significant associations emerged with positive health interaction (β=.238, *P*=.001), health communication (β=.350, *P*=.001), and nonadherence (β=.206, *P*=.001). There were no significant associations with strained health interaction in the full sample or subsample.



XSL•FO
RenderX

**Table 4.** Ordinary least regression analysis of the e-HLS factorial structure for the full sample (n=710).

| | Health interaction | | Health communication | | Strain | | Empowerment | | Negative affect | | Health problem | | Nonadherence | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| e-HLS Factors | *b* | β (*P*) | *b* | β (*P*) | *b* | β (*P*) | *b* | β (*P*) | *b* | β (*P*) | *b* | β (*P*) | *b* | β (*P*) |
| Action | 0.055 | .077 (.06) | 0.163 | .206 (.001) | -0.027 | -.033 (.46) | 0.254 | .303 (.001) | -0.153 | -.174 (.001) | 0.013 | .032 (.49) | 0.193 | .316 (.001) |
| Communication | .235 | .302 (.001) | 0.279 | .323 (.001) | 0.017 | .019 (.68) | 0.097 | .106 (.001) | 0.072 | .075 (.11) | 0.067 | .147 (.002) | 0.128 | .191 (.001) |
| Trust | .308 | .282 (.001) | 0.340 | .280 (.001) | 0.116 | .092 (.02) | 0.377 | .293 (.001) | 0.077 | .077 (.04) | -0.007 | -.011 (.78) | 0.204 | .216 (.001) |
| $R^2$ | | .247 | | .359 | | .009 | | .266 | | .025 | | .012 | | .292 |
| Adjusted $R^2$ | | .243 | | .356 | | .005 | | .263 | | .021 | | .023 | | .289 |

**Table 5.** Ordinary least regression analysis of the e-HLS factorial structure for the subsample (n=194).

| | Health interaction | | Health communication | | Strain | | Empowerment | | Negative affect | | Health problem | | Nonadherence | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| e-HLS Factors | *b* | β (*P*) | *b* | β (*P*) | *b* | β (*P*) | *b* | β (*P*) | *b* | β (*P*) | *b* | β (*P*) | *b* | β (*P*) |
| Action | 0.016 | .024 (.77) | 0.098 | .140 (.05) | -0.121 | -.165 (.07) | 0.273 | .365 (.001) | -0.191 | -.250 (.005) | 0.027 | .109 (.24) | 0.187 | .312 (.001) |
| Communication | 0.193 | .238 (.004) | 0.296 | .350 (.001) | 0.015 | .017 (.85) | 0.053 | .059 (.45) | 0.072 | .078 (.38) | 0.002 | .006 (.95) | 0.149 | .206 (.007) |
| Trust | 0.364 | .340 (.001) | 0.364 | .326 (.001) | -0.089 | -.075 (.31) | 0.356 | .299 (.001) | 0.137 | .112 (.13) | -0.034 | -.088 (.24) | 0.239 | .249 (.001) |
| $R^2$ | | .220 | | .375 | | .033 | | .298 | | .052 | | .017 | | .329 |
| Adjusted $R^2$ | | .207 | | .365 | | .018 | | .286 | | .037 | | .001 | | .319 |



XSL•FO
RenderX

**Figure 1.** Confirmatory Factor Analysis of the e-HLS Items (n=710).

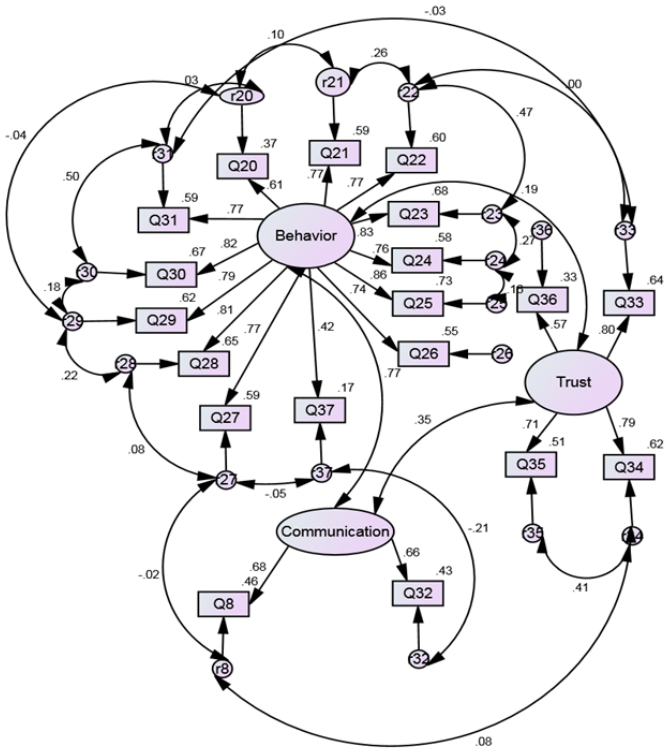
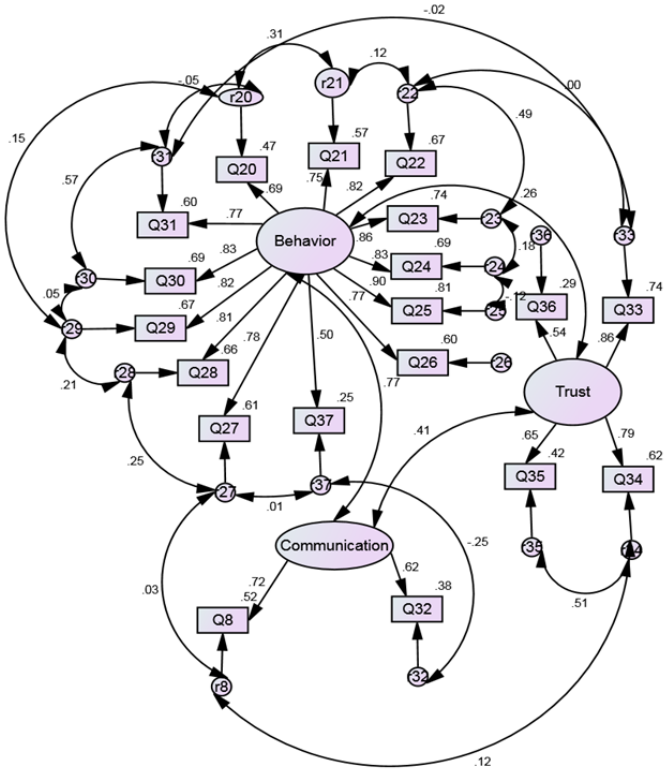



**Figure 2.** Confirmatory Factor Analysis of the e-HLS Items (n=194).






XSL•FO
RenderX

## Discussion

The Institute of Medicine's recommendation to expand the scope of health literacy by considering multiple skills has led to increased recognition that a comprehensive examination of health literacy in the digital environment is needed [18,66-75]. Our research study is also consistent with the National Call to Action to Promote Health Literacy, which suggested the need to develop multidimensional measures of health literacy to include skills beyond the comprehension of written health information [76].

The strength of our measure is its contemporary multidimensional view of health literacy. We expand its conceptualization beyond the traditional document-based measures (being able to find and understand information) to include interactive and communicative aspects of literacy (information exchange) and critical evaluative skills of information (quality assessment) provided in electronic sources. We created our instrument to comprise of 3 domains: behavioral literacy (action factor), cognitive literacy (trust factor), and interactional literacy (communication factor). Thus, our measure expands the understanding of ehealth literacy through the addition of the 3 domains. Because the concept of health literacy is increasingly conceptualized as consisting of skills related to evaluating, communicating, and using information to make informed decisions, we designed our new measure to reflect these skills.

The results provide statistical support for the multidimensionality of our scale. Consistent with recent studies, our scale suggest that ehealth literacy includes a broader array of skills besides the ability to read and understand health information stressing the need to focus on multiple dimensions of content areas and skills [30,33,77]. The ability of this composite scale to provide information about the extent to which people assess the quality and credibility of ehealth information makes it a valuable assessment tool. Although the composite scale yields a single score, combining data from items that are loaded onto separate factors into a single score may suppress potential differences that can be found if the scale factors are analyzed separately. In our study, we first evaluated the psychometric properties of a new ehealth literacy measure in a national sample of Internet users. Confirmatory factor analysis was conducted to determine whether the 3-factor structure of the e-HLS, as suggested by EFA, achieved a good fit with our entire sample and subsample. The goodness-of-fit indices provided by CFA confirmed the robustness of the e-HLS. Each of the scale factors demonstrated a good internal consistency and validity. We examined patterns of correlations of our measure with related covariates. Moreover, we regressed scale factors on variables that literature has shown to be associated with health literacy in the general sample [51].These analytical approaches further validated our new measure. However, some limitations in our study must be acknowledged. First, we had to rely on self-reported cross-sectional data, and we lacked useful information illustrating the extent to which our respondents used the Internet for health information. Furthermore, we were not able to measure health literacy to its fullest dimensions, as a wide range of skills and behaviors comprise health literacy. Moreover, as a means of evaluating health literacy, self-report may not be entirely accurate. Regardless, this method may further improve our understanding of the role of health literacy in the daily lives of Americans and provide constructive information to that end.

As summarized in Table 1, the fact that our respondents reported high confidence in discerning information quality and trust in Internet information while reporting low rates of behavior to verify information credibility and quality suggests low awareness about the questionable trustworthiness and credibility of information found on the Internet. Communication with health professionals for purposes of asking advice about which websites they should consult and where to find credible information on the Internet was not common either. As patients want a greater understanding and more active role in in their health management, we need new and expanded approaches to examine health literacy to incorporate the provision of credible information through the new digital technologies [77,78]. When examining indicators of electronic health literacy among our respondents, we found higher scores among those with higher levels of education and income. The factorial structure of our scale explained the greatest variance for respondents' perceived positive changes in health communication with their providers followed by patient nonadherence, perceived empowerment, and positive health interaction. On the other hand, the factor structure of our scale had little explanatory power for perceived strain in health interactions, negative effect, and experience of a health problem. Our results suggest that satisfaction with medical encounters is enhanced by consumers with higher levels of ehealth literacy.

Examination of standardized beta coefficients revealed that the trust factor of our measure had a strong association with perceived empowerment, suggesting that those respondents who trusted information gathered from the Internet to a greater extent also reported higher sense of empowerment. Consistent with existing research, around 60% of our respondents reported relatively little skepticism of the quality of Internet health information despite the fact that much of the Internet health or medical information is of questionable accuracy and lacks any endorsement or sanction of by a formal medical authority [63,70,79]. This is particularly surprising given that less than half of our participants reported having performed a quality check of the Internet information on most of our scale items. We found that they trusted information provided by health websites, and they overestimated the credibility and accuracy of information presented. This is despite various challenges associated with the Internet search, including information overload, navigating through hundreds of search results, many of which could be irrelevant, and separating questionable from credible health information. Moreover, face-to-face, traditional medical encounters may become awkward or strained now that health care professionals no longer enjoy an information monopoly. The way patients can engage with health information, including accessing information, about unverified alternative medicines, may pose challenges for doctors and other providers who need to interact personally with consumers who can now gather their information from the Internet. This actually may form the crux of the problems troubling health professionals;



XSL•FO
RenderX

patients exercise their elevated sense of health care empowerment yet remain insufficiently cognizant of the real dangers of trying to manage their health based on inconsistent and potentially highly inaccurate Web-based information. Health care professionals' fears may be well founded. Our regression analyses indicated that information seekers who place a great deal of trust in Internet information report greater levels of nonadherence. In addition, the trust in Internet information (trust factor) is significantly associated with patient nonadherence to doctor's guidelines and/or treatments, further confirming professionals' concerns.

Not surprisingly, a high level of trust in Web-based information had a significant positive association with strained interactions during medical encounters as reported by our research participants. Interestingly, although we found a significant association between trust in the Internet information and reports of perceived strain in medical encounters in our full sample, this association was not significant in the subsample. This might suggest that health professionals were less likely to feel challenged or distrusted by older patients using the Internet to seek health or medical information about their concerns than younger patients. In addition, the action factor had a negative significant association with negative effect (worry and/or anxiety), whereas the trust in Internet factor had a positive association with negative effect. These findings suggest that respondents who took action to evaluate health information reported less worry and/or anxiety, whereas those who placed a great deal of trust in the Internet information reported more worry and/or anxiety.

The action factor of our scale explained the most variance in perceived empowerment and nonadherence. Respondents who engaged in various quality checks of health information seem to perceive themselves as better equipped to cope with their health concern or issue. On the other hand, the communication factor had the highest explanatory power for positive health interaction and health communication. These associations suggest that health consumers who work with their health care professionals to find the most credible sources before they search the Internet perceive positive changes in their encounters with their providers. When patients share the information they discovered on the Web with their providers, they ask more informed questions and better understand the doctor's information. Moreover, they perceive respect from their providers as partners in the health care process.

In contrast, perceived sense of empowerment, as a result of information obtained from the Internet sources, without communication with a health care provider is associated with increased rates of noncompliance with treatment and medical advice of a health professional. Accordingly, further examination of sociotechnological changes and their effect on doctor–patient interaction and communication is warranted.

## Conflicts of Interest

None declared.

## References


1. Brown N, Webster A. New medical technologies and society: reordering life. Hoboken, NJ: Polity; 2004.
2. Beaudoin CE, Tao CC. Benefiting from social capital in online support groups: an empirical study of cancer patients. Cyberpsychol Behav 2007 Aug;10(4):587-590. [doi: 10.1089/cpb.2007.9986] [Medline: 17711369]
3. Ratzan SC. Health literacy: building upon a strong foundation. J Health Commun 2011;16 Suppl 3:3-4. [doi: 10.1080/10810730.2011.606071] [Medline: 21951237]
4. Fox S, Duggan M. Pew Internet. Health Online 2013 URL: http://www.pewinternet.org/2013/01/15/health-online-2013/ [accessed 2016-06-03] [WebCite Cache ID 6hz1pIXUz]
5. Wen LM, Rissel C, Baur LA, Lee E, Simpson JM. Who is NOT likely to access the Internet for health information? Findings from first-time mothers in southwest Sydney, Australia. Int J Med Inform 2011 Jun;80(6):406-411. [doi: 10.1016/j.ijmedinf.2011.03.001] [Medline: 21493126]
6. Ratzan SC. Web 2.0 and health communication. J Health Commun 2011;16 Suppl 1:1-2. [doi: 10.1080/10810730.2011.601967] [Medline: 21843091]
7. Atienza AA, Hesse BW, Gustafson DH, Croyle RT. E-health research and patient-centered care examining theory, methods, and application. Am J Prev Med 2010 Jan;38(1):85-88. [doi: 10.1016/j.amepre.2009.10.027] [Medline: 20117562]
8. Lupiáñez-Villanueva F, Hardey M, Torrent J, Ficapal P. The integration of Information and Communication Technology into medical practice. Int J Med Inform 2010 Jul;79(7):478-491. [doi: 10.1016/j.ijmedinf.2010.04.004] [Medline: 20472494]
9. Schulz PJ, Nakamoto K. Health literacy and patient empowerment in health communication: the importance of separating conjoined twins. Patient Educ Couns 2013 Jan;90(1):4-11. [doi: 10.1016/j.pec.2012.09.006] [Medline: 23063359]
10. Dutta-Bergman M. Media use theoryinternet use for health care. In: Murero M, Rice RE, editors. The internet and health care: theory, research, and practice. Mahwah, N.J: Larence Erlbaum Associates; 2006:83-103.
11. Aymé S, Kole A, Groft S. Empowerment of patients: lessons from the rare diseases community. Lancet 2008 Jun 14;371(9629):2048-2051. [doi: 10.1016/S0140-6736(08)60875-2] [Medline: 18555918]
12. Seckin G. Digitized emotions in the pocket: social computing and mobile mental health (mMH) applications. In: Tettegah SY, Garcia EY, editors. Emotions, technology, and health. San Diego, California: Academic Press; 2016:41-61.
13. Luciano JS, Cumming GP, Wilkinson MD, Kahana E. The emergent discipline of health web science. J Med Internet Res 2013;15(8):e166 [FREE Full text] [doi: 10.2196/jmir.2499] [Medline: 23968998]




XSL•FO
RenderX

14. Lober WB, Flowers JL. Consumer empowerment in health care amid the internet and social media. Semin Oncol Nurs 2011 Aug;27(3):169-182. [doi: 10.1016/j.soncn.2011.04.002] [Medline: 21783008]
15. Seidman J. The mysterious maze of the world wide web: How can we guide consumers to high-quality health information on the internet? In Murero M, Rice RE. In: The Internet and Health Care. Theory, Research, and Practice. Mahwah, New Jersey: Lawrence-Erlbaum Associates.. ISBN. Mahwah, New Jersey: Lawrence-Erlbaum; 2006:195-212.
16. Dedding C, van DR, Winkler L, Reis R. How will e-health affect patient participation in the clinic? A review of e-health studies and the current evidence for changes in the relationship between medical professionals and patients. Soc Sci Med 2011 Jan;72(1):49-53. [doi: 10.1016/j.socscimed.2010.10.017] [Medline: 21129832]
17. Weaver JB, Thompson NJ, Weaver SS, Hopkins GL. Healthcare non-adherence decisions and internet health information. Computers in Human Behavior 2009 Nov;25(6):1373-1380. [doi: 10.1016/j.chb.2009.05.011]
18. Nielsen-Bohlman L, Panzer A, Kindig D. Health Literacy: A Prescription to End Confusion. Washington, DC: The National Academies Press; 2004.
19. Seçkin G. Health information on the web and consumers' perspectives on health professionals' responses to information exchange. Med 2 0 2014;3(2):e4 [FREE Full text] [doi: 10.2196/med20.3213] [Medline: 25075248]
20. Im E. Online support of patients and survivors of cancer. Semin Oncol Nurs 2011 Aug;27(3):229-236. [doi: 10.1016/j.soncn.2011.04.008] [Medline: 21783014]
21. Bickmore T, Paasche-Orlow M. The role of information technology in health literacy research. J Health Commun. (Suppl. 3) 2013;17:23-29. [Medline: 3030559]
22. Health literacy and consumer-facing technology: Workshop summary. Washington, D.C: National Academies of Science, Engineering, and Medicine; 2015.
23. Seçkin G. Smart phone health applications. In: Kahana E, editor. The Encyclopedia of Mobile Phone Behavior. Hershey, Pennsylvania: Information Science Reference; 2015:898-905.
24. Pleasant A, McKinney J, Rikard RV. Health literacy measurement: a proposed research agenda. J Health Commun 2011;16 Suppl 3:11-21. [doi: 10.1080/10810730.2011.604392] [Medline: 21951240]
25. Guzys D, Kenny A, Dickson-Swift V, Threlkeld G. A critical review of population health literacy assessment. BMC Public Health 2015;15:215 [FREE Full text] [doi: 10.1186/s12889-015-1551-6] [Medline: 25885742]
26. Sørensen K, Van den Broucke Stephan, Fullam J, Doyle G, Pelikan J, Slonska Z, (HLS-EU) Consortium Health Literacy Project European. Health literacy and public health: a systematic review and integration of definitions and models. BMC Public Health 2012;12:80 [FREE Full text] [doi: 10.1186/1471-2458-12-80] [Medline: 22276600]
27. Paasche-Orlow MK, Wilson Elizabeth A H, McCormack L. The evolving field of health literacy research. J Health Commun 2010;15 Suppl 2:5-8. [doi: 10.1080/10810730.2010.499995] [Medline: 20845188]
28. Haun J, Luther S, Dodd V, Donaldson P. Measurement variation across health literacy assessments: implications for assessment selection in research and practice. J Health Commun 2012;17 Suppl 3:141-159. [doi: 10.1080/10810730.2012.712615] [Medline: 23030567]
29. Dumenci L, Matsuyama R, Riddle DL, Cartwright LA, Perera RA, Chung H, et al. Measurement of cancer health literacy and identification of patients with limited cancer health literacy. J Health Commun 2014;19 Suppl 2:205-224 [FREE Full text] [doi: 10.1080/10810730.2014.943377] [Medline: 25315594]
30. Coleman CA, Hudson S, Maine LL. Health literacy practices and educational competencies for health professionals: a consensus study. J Health Commun 2013;18 Suppl 1:82-102 [FREE Full text] [doi: 10.1080/10810730.2013.829538] [Medline: 24093348]
31. Rubin DL, Parmer J, Freimuth V, Kaley T, Okundaye M. Associations between older adults' spoken interactive health literacy and selected health care and health communication outcomes. J Health Commun 2011;16 Suppl 3:191-204. [doi: 10.1080/10810730.2011.604380] [Medline: 21951252]
32. Tennant B, Stellefson M, Dodd V, Chaney B, Chaney D, Paige S, et al. eHealth literacy and Web 2.0 health information seeking behaviors among baby boomers and older adults. J Med Internet Res 2015;17(3):e70 [FREE Full text] [doi: 10.2196/jmir.3992] [Medline: 25783036]
33. Zarcadoolas C, Pleasant A, Greer D. Advancing health literacy: A framework for understanding and action. San Francisco, California: Jossey-Bass; 2006.
34. Pleasant A. Advancing health literacy measurement: a pathway to better health and health system performance. J Health Commun 2014 Dec;19(12):1481-1496 [FREE Full text] [doi: 10.1080/10810730.2014.954083] [Medline: 25491583]
35. Coleman C, Kurtz-Rossi S, McKinney J, Pleasant A, Rootman I. The Calgary Charter on Health Literacy: Rationale and Core Principles for the Development of Health Literacy Curricula. Calgary, Montreal: The Center for Literacy; 2008.
36. Norman C. eHealth literacy 2.0: problems and opportunities with an evolving concept. J Med Internet Res 2011;13(4):e125 [FREE Full text] [doi: 10.2196/jmir.2035] [Medline: 22193243]
37. Norman CD, Skinner HA. eHealth Literacy: Essential Skills for Consumer Health in a Networked World. J Med Internet Res 2006 Jun;8(2):e9 [FREE Full text] [doi: 10.2196/jmir.8.2.e9] [Medline: 16867972]
38. Norgaard O, Furstrand D, Klokker L, Karnoe A, Batterham R, Kayser L, et al. The e-health literacy framework: A conceptual framework for characterizing e-health users and their interaction with e-health systems. Know Man E-Learn 2015;7(4):522-540.




XSL•FO
RenderX

39. Schonlau M, Martin L, Haas A, Derose KP, Rudd R. Patients' literacy skills: more than just reading ability. J Health Commun 2011 Nov;16(10):1046-1054 [FREE Full text] [doi: 10.1080/10810730.2011.571345] [Medline: 21916699]
40. Murero M, Rice RE. e-health: beyond the internet searches. In: The internet and health care: theory, research, and practice. Mahwah, NJ: Larence Erlbaum Associates; 2006:29-48.
41. Norman CD, Skinner HA. eHEALS: The eHealth Literacy Scale. J Med Internet Res 2006 Nov;8(4):e27 [FREE Full text] [doi: 10.2196/jmir.8.4.e27] [Medline: 17213046]
42. Jordan JE, Osborne RH, Buchbinder R. Critical appraisal of health literacy indices revealed variable underlying constructs, narrow content and psychometric weaknesses. J Clin Epidemiol 2011 Apr;64(4):366-379. [doi: 10.1016/j.jclinepi.2010.04.005] [Medline: 20638235]
43. Chew LD, Bradley KA, Boyko EJ. Brief questions to identify patients with inadequate health literacy. Fam Med 2004 Sep;36(8):588-594 [FREE Full text] [Medline: 15343421]
44. Hargittai E, Hsieh YP. Succinct Survey Measures of Web-Use Skills. Social Science Computer Review 2011 Feb 28;30(1):95-107. [doi: 10.1177/0894439310397146]
45. Koopman RJ, Petroski GF, Canfield SM, Stuppy JA, Mehr DR. Development of the PRE-HIT instrument: patient readiness to engage in health information technology. BMC Fam Pract 2014;15:18 [FREE Full text] [doi: 10.1186/1471-2296-15-18] [Medline: 24472182]
46. Manafò E, Wong S. eSEARCH©: A Tool to Promote the eHealth Literacy Skills of Older Adults. Journal of Consumer Health On the Internet 2013 Jul;17(3):255-271. [doi: 10.1080/15398285.2013.812914]
47. Murphy P, Davis T, Long S, Jackson R, Decker B. Rapid estimate of adult literacy in medicine (REALM): A quick reading test for patients. J Reading 1993;37(2):124-130.
48. Karnoe A, Kayser L. How is eHealth literacy measured and what do the measurements tell us? A systematic review. Know Manage E-Learn 2015;7(4):576-600.
49. Osborne RH, Batterham RW, Elsworth GR, Hawkins M, Buchbinder R. The grounded psychometric development and initial validation of the Health Literacy Questionnaire (HLQ). BMC Public Health 2013;13:658 [FREE Full text] [doi: 10.1186/1471-2458-13-658] [Medline: 23855504]
50. van der Vaart Rosalie, van Deursen Alexander Jam, Drossaert CH, Taal E, van Dijk Jan Amg, van de Laar Mart Afj. Does the eHealth Literacy Scale (eHEALS) measure what it intends to measure? Validation of a Dutch version of the eHEALS in two adult populations. J Med Internet Res 2011;13(4):e86 [FREE Full text] [doi: 10.2196/jmir.1840] [Medline: 22071338]
51. Cafiero M. Nurse practitioners' knowledge, experience, and intention to use health literacy strategies in clinical practice. J Health Commun 2013;18 Suppl 1:70-81 [FREE Full text] [doi: 10.1080/10810730.2013.825665] [Medline: 24093347]
52. Knowledge N. Knowledge Panel Design Summary. 2013. URL: http://www.knowledgenetworks.com/knpanel/docs/knowledgepanel(R)-design-summary-description.pdf [accessed 2016-06-03] [WebCite Cache ID 6hzI8TRO3]
53. Hair J, Black W, Babin B, Anderson R, Tatham R. Multivriate Data Analysis. Upper Saddle, New Jersey: Prentice Hall; 2006.
54. Kesici S, Sahin I. A comparative study of uses of the Internet among college students with and without Internet addiction. Psychol Rep 2009 Dec;105(3 Pt 2):1103-1112. [doi: 10.2466/PR0.105.F.1103-1112] [Medline: 20229914]
55. Khazaal Y, Billieux J, Thorens G, Khan R, Louati Y, Scarlatti E, Van Der Linden Martial, et al. French validation of the internet addiction test. Cyberpsychol Behav 2008 Dec;11(6):703-706. [doi: 10.1089/cpb.2007.0249] [Medline: 18954279]
56. Spector P. Summated rating scale construction. Newbury Park, California: Sage; 1992.
57. Ishikawa H, Takeuchi T, Yano E. Measuring functional, communicative, and critical health literacy among diabetic patients. Diabetes Care 2008 May;31(5):874-879. [doi: 10.2337/dc07-1932] [Medline: 18299446]
58. A report on the quality of health information on the Internet. Oakland, CA: California Health Care Foundation; 2001.
59. Baumgartner SE, Hartmann T. The role of health anxiety in online health information search. Cyberpsychol Behav Soc Netw 2011 Oct;14(10):613-618. [doi: 10.1089/cyber.2010.0425] [Medline: 21548797]
60. Briggs P, Burford B, De Angeli A, Lynch P. Trust in Online Advice. Social Science Computer Review 2002 Aug 01;20(3):321-332. [doi: 10.1177/089443930202000309]
61. Hu Y, Shyam Sundar S. Effects of Online Health Sources on Credibility and Behavioral Intentions. Communication Research 2009 Nov 25;37(1):105-132. [doi: 10.1177/0093650209351512]
62. Hong T. Internet health information in the patient-provider dialogue. Cyberpsychol Behav 2008 Oct;11(5):587-589. [doi: 10.1089/cpb.2007.0172] [Medline: 18771392]
63. Vanderminden J, Potter S. Challenges to the Doctor-Patient Relationship in the Twenty-First Century. In: Cockerham WC, editor. The New Blackwell Companion to Medical Sociology. Malden, MA: Blackwell Publishing; 2010:355-372.
64. Morahan-Martin JM. How internet users find, evaluate, and use online health information: a cross-cultural review. Cyberpsychol Behav 2004 Oct;7(5):497-510. [Medline: 15667044]
65. Eysenbach G, Powell J, Kuss O, Sa ER. Empirical studies assessing the quality of health information for consumers on the world wide web: a systematic review. JAMA 2002;287(20):2691-2700. [Medline: 12020305]
66. Murero M, Rice RE. Patients? Online Information-Seeking Behavior. In: The internet and health care: theory, research, and practice. Mahwah, NJ: Larence Erlbaum Associates; 2006:127-149.




XSL•FO
RenderX

67. U.S. National Library of Medicine. 2016. Evaluating health information URL: https://www.nlm.nih.gov/medlineplus/evaluatinghealthinformation.html [accessed 2016-06-02] [WebCite Cache ID 6hxRoVwoc]
68. Peterson G, Aslani P, Williams K. How do consumers search for and appraise information on medicines on the Internet? A qualitative study using focus groups. J Med Internet Res 2003 Dec 19;5(4):e33 [FREE Full text] [doi: 10.2196/jmir.5.4.e33] [Medline: 14713661]
69. Ybarra M, Suman M. Reasons, assessments and actions taken: sex and age differences in uses of Internet health information. Health Educ Res 2008 Jun;23(3):512-521 [FREE Full text] [doi: 10.1093/her/cyl062] [Medline: 16880222]
70. Seçkin G. Informational and decisional empowerment in online health support communities: initial psychometric validation of the Cyber Info-Decisional Empowerment Scale (CIDES) and preliminary data from administration of the scale. Support Care Cancer 2011 Dec;19(12):2057-2061. [doi: 10.1007/s00520-011-1249-y] [Medline: 21858718]
71. DeCoster J. Stat-help. 2005. Scale construction notes URL: http://www.stat-help.com/scale.pdf [accessed 2016-06-03] [WebCite Cache ID 6hzIykeUB]
72. Todman J. Dugard, Approaching Multivariate Analysis. In: An Introduction for Psychology. New York: Psychology Press; 2007.
73. Kenny D, Kashy D, Cook WL, Simpson JA. Dyadic data analysis. In: Dyadic Data Analysis (Methodology In The Social Sciences). New York: The Guilford Press; 2006.
74. Berkman ND, Sheridan SL, Donahue KE, Halpern DJ, Crotty K. Low health literacy and health outcomes: an updated systematic review. Ann Intern Med 2011 Jul 19;155(2):97-107. [doi: 10.7326/0003-4819-155-2-201107190-00005] [Medline: 21768583]
75. Berkman ND, Davis TC, McCormack L. Health literacy: what is it? J Health Commun 2010;15 Suppl 2:9-19. [doi: 10.1080/10810730.2010.499985] [Medline: 20845189]
76. U.S. Department of HealthHuman Services, Office of Disease PreventionHealth Promotion. National Action Plan to Improve Health Literacy. Washington, D.C: U.S. Department of Health and Human Services; 2010.
77. Cutilli CC, Bennett I. Understanding the Health Literacy of America. Orthopaedic Nursing 2009;28(1):27-32. [doi: 10.1097/01.NOR.0000345852.22122.d6]
78. Murero M, Rice RE. The Internet and health care. In: Theory, research, and practice. Mahwah, New Jersey: Lawrence-Erlbaum Associates; 2006.
79. Beck F, Richard J, Nguyen-Thanh V, Montagni I, Parizot I, Renahy E. Use of the internet as a health information resource among French young adults: results from a nationally representative survey. J Med Internet Res 2014;16(5):e128 [FREE Full text] [doi: 10.2196/jmir.2934] [Medline: 24824164]


## Abbreviations

**BTS:** Bartlett Test of Sphericity
**CFA:** confirmatory factor analysis
**CFI:** comparative fit index
**EFA:** exploratory factor analysis
**e-HLS:** electronic health literacy
**KMO:** Kaiser–Meyer–Olkin
**KN:** Knowledge Networks
**NFI:** normed fit index
**RMSEA:** root mean squared error of approximation


*Edited by A Moorhead; submitted 02.01.16; peer-reviewed by I Montagni, P Schulz; comments to author 21.02.16; revised version received 01.05.16; accepted 05.05.16; published 11.07.16.*

*Please cite as:*
*Seçkin G, Yeatts D, Hughes S, Hudson C, Bell V*
*Being an Informed Consumer of Health Information and Assessment of Electronic Health Literacy in a National Sample of Internet Users: Validity and Reliability of the e-HLS Instrument*
*J Med Internet Res 2016;18(7):e161*
*URL: http://www.jmir.org/2016/7/e161/*
*doi:10.2196/jmir.5496*
*PMID:27400726*


©Gül Seçkin, Dale Yeatts, Susan Hughes, Cassie Hudson, Valarie Bell. Originally published in the Journal of Medical Internet Research (http://www.jmir.org), 11.07.2016. This is an open-access article distributed under the terms of the Creative Commons



XSL•FO
RenderX

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



XSL•FO
RenderX

Original Paper

# Associations of eHealth Literacy With Health Behavior Among Adult Internet Users

Seigo Mitsutake[1], PT, PhD; Ai Shibata[2], PhD; Kaori Ishii[3], PhD; Koichiro Oka[3], PhD

[1]Research Team for Human Care, Tokyo Metropolitan Institute of Gerontology, Tokyo, Japan

[2]Faculty of Health and Sport Sciences, University of Tsukuba, Ibaraki, Japan

[3]Faculty of Sport Sciences, Waseda University, Saitama, Japan

**Corresponding Author:**
Seigo Mitsutake, PT, PhD
Research Team for Human Care
Tokyo Metropolitan Institute of Gerontology
35-2 Sakae-cho, Itabashi-ku
Tokyo,
Japan
Phone: 81 3 3964 3241 ext 4229
Fax: 81 3 3579 4776
Email: mitsu@tmig.or.jp

## Abstract

**Background:** In the rapidly developing use of the Internet in society, eHealth literacy—having the skills to utilize health information on the Internet—has become an important prerequisite for promoting healthy behavior. However, little is known about whether eHealth literacy is associated with health behavior in a representative sample of adult Internet users.

**Objective:** The aim of this study was to examine the association between eHealth literacy and general health behavior (cigarette smoking, physical exercise, alcohol consumption, sleeping hours, eating breakfast, eating between meals, and balanced nutrition) among adult Internet users in Japan.

**Methods:** The participants were recruited among registrants of a Japanese Internet research service company and asked to answer a cross-sectional Internet-based survey in 2012. The potential respondents (N=10,178) were randomly and blindly invited via email from the registrants in accordance with the set sample size and other attributes. eHealth literacy was assessed using the Japanese version of the eHealth Literacy Scale. The self-reported health behaviors investigated included never smoking cigarettes, physical exercise, alcohol consumption, sleeping hours, eating breakfast, not eating between meals, and balanced nutrition. We obtained details of sociodemographic attributes (sex, age, marital status, educational attainment, and household income level) and frequency of conducting Internet searches. To determine the association of each health behavior with eHealth literacy, we performed a logistic regression analysis; we adjusted for sociodemographic attributes and frequency of Internet searching as well as for other health behaviors that were statistically significant with respect to eHealth literacy in univariate analyses.

**Results:** We analyzed the data of 2115 adults (response rate: 24.04%, 2142/10,178; male: 49.74%, 1052/2115; age: mean 39.7, SD 10.9 years) who responded to the survey. Logistic regression analysis showed that individuals with high eHealth literacy were significantly more likely to exhibit the good health behaviors of physical exercise (adjusted odds ratio [AOR] 1.377, 95% CI 1.131-1.678) and eating a balanced diet (AOR 1.572, 95% CI 1.274-1.940) than individuals with low eHealth literacy.

**Conclusions:** We found that some health behaviors, including exercise and balanced nutrition, were independently associated with eHealth literacy among Japanese adult Internet users.

***(J Med Internet Res 2016;18(7):e192)*** doi:10.2196/jmir.5413

**KEYWORDS**

health literacy; ehealth literacy; epatients; Internet; health behavior; cross-sectional studies



XSL•FO
RenderX

## Introduction

According to an estimate of the Communications Usage Trend Survey in 2013, 82.8% of Japan's general population are Internet users [1]. Approximately 70% of Japanese Internet users seek health information online [2]. One US study indicated that 72% of Internet users had looked online for health information over the previous year [3]; 59% of those who looked online for health information did so specifically to determine what medical condition they or an acquaintance might have [3]. In addition to improved medication compliance, decreased anxiety, and a greater feeling of safety, Internet users exhibit better self-care health behavior than those who do not use the Internet [4,5]. Thus, the Internet is increasingly becoming an effective information tool for attaining and maintaining better self-care health behavior [6,7].

In an information society, health literacy is growing in importance with respect to public health, and health care involves effectively using health information from multiple sources [8,9]. Health literacy—the degree to which individuals can obtain, process, and understand basic health information and services needed to make appropriate health decisions—is a key competence in promoting individual and public health [9]. Previous studies have identified an association between low health literacy and decreased knowledge of health care services and self-care management skills [8,10]. Toward improving health care quality and population health outcomes and achieving health equity, promoting health literacy is indicated as one of the objects of health communication and health information technology in "Healthy People 2020″ [11].

In this context, health information primarily relates to such electronic resources as the Internet and other technologies. Health information has notably assumed an important role in health promotion among the general public through the widespread use of personal computers and smartphones/mobile phones [6,7,12]. With this proliferation of online health information, one critical issue to have emerged is that many websites providing health information are invalid or difficult to understand; they may also be linked to commercial goods or private health services [13-15]. Regulating health information on the Internet is difficult because new information is constantly added. To utilize health information on the Internet properly, people seeking such information need to obtain "the ability to seek, find, understand, and appraise health information from electronic sources and apply the knowledge gained to addressing or solving a health problem" (ie, eHealth literacy) [15].

Previous studies of eHealth literacy have largely focused on defining the term [15-18], developing measures of eHealth literacy [2,19-23], and examining the effect of eHealth literacy interventions on people in need of it [24-27]. More recently, studies on eHealth literacy have examined the association between eHealth literacy and health-related outcomes. The Integrative Model of eHealth Use (IMeHU) suggests that social structures affect health behaviors through the microlevel conditions of eHealth literacy, motivation, and efficacy in using the Internet for health purposes [28]. Empirical studies have shown that individuals with high eHealth literacy had greater efficacy in finding health information and using health apps [29,30], were more active health information seekers [31-33], and employed more search strategies [33,34] than people with low eHealth literacy. Moreover, a limited number of studies have identified an association between eHealth literacy and health behaviors [35,36]. Hsu et al [36] showed that eHealth literacy mediated the association between individual factors and health behavior among college students; therefore, promoting health behavior among such students demands high levels of eHealth literacy. However, few studies have examined the relationship between eHealth literacy and health behavior in a general population.

The aim of the Healthy Japan 21 (second term) campaign of Japan's Ministry of Health, Labour and Welfare is to prevent chronic diseases and improve daily health behavior with respect to smoking, exercise, alcohol, rest, and dietary habits among Japanese adults [37]. As in other developed countries, the Internet in Japan is a powerful means of promoting healthy behavior among college students as well as the adult population [1,2]. To design effective strategies for promoting healthy behavior among adult Internet users, it is necessary to examine the relationship between Internet use and such behavior. According to the IMeHU, eHealth literacy may play an important role in health behavior; however, little is known about the precise association in an adult population. Therefore, this study examines whether eHealth literacy is associated with various kinds of general health behavior: cigarette smoking, physical exercise, alcohol consumption, sleeping hours, eating breakfast, eating between meals, and balanced nutrition.

## Methods

### Participants

The study participants were recruited from the registrants of a Japanese Internet research service company called MyVoice Communication, Inc; the recipients were asked to respond to a cross-sectional Internet-based survey in 2012. In this study, we recruited individual Internet users because eHealth literacy is necessary to access online health information. We believed that an Internet survey would be appropriate for this study because responders to such a survey are clearly able to use the Internet effectively. The research company had approximately 1,180,000 voluntarily registered participants in 2012, and it obtained detailed sociodemographic data from each participant on registration. In this study, we aimed to collect data from 2000 men and women aged 20 to 59 years. We intended to minimize selection bias caused by proportional differences in terms of sex and age; therefore, we allocated the registered participants equally to eight sample groups categorized by sex and age (20-29, 30-39, 40-49, and 50-59 years), with n=250 in each group. The Internet research service company randomly chose the potential respondents from the registered participants in accordance with the sample sizes: N=10,178; male: 20-29 years, n=2275; 30-39 years, n=1255; 40-49 years, n=880; 50-59 years, n=699; female: 20-29 years, n=1979; 30-39 years, n=1362; 40-49 years, n=963; and 50-59 years, n=765. In addition, the Internet research service company blindly selected the potential



XSL•FO
RenderX

respondents such that the authors and other registered participants were unable to identify those individuals.

The company invited registrants to participate in the survey by email. The number of potential respondents in each stratified sample group was determined by dividing the quota (n=250) by the response rate for the corresponding sociodemographic group. That response rate was computed based on the results of many previous surveys conducted by the research company. The questionnaires were placed in a protected area of a website, and the potential respondents received a specific URL in their invitation email. Potential respondents were able to log on to the protected area of the site using a unique ID and password. After the desired number of participants had voluntarily signed an online informed consent form and completed the sociodemographic data information form, further participants were no longer accepted. Reward points valued at 150 yen were provided as incentives for participation (US $1 was equivalent to approximately 82 yen in 2012). This study was approved by the Ethics Committee of Waseda University, Tokyo, Japan (No: 2011-245).

## Measurements

### *Sociodemographic Attributes*

The research company provided categorized data as follows: sex (male, female); age group (20-29, 30-39, 40-49, and 50-59 years); marital status (not married, married); education level (up to high school, 2-year college or career college, college graduate or above); and household income level (<5 million yen, ≥5 million yen).

### *Frequency of Internet Searching*

We assessed the frequency of information searches on the Internet in terms of daily conducted searches. We did so because one study found a positive association between eHealth literacy and the frequency of Internet searches [2]; we believed that the frequency of Internet searching could be used as a control variable for eHealth literacy and healthy behavior.

### *eHealth Literacy*

We used the Japanese version of the eHealth Literacy Scale (eHEALS) to assess the eHealth literacy levels of participants [2]. The eHEALS consists of eight questions (see Multimedia Appendix 1); it uses a 5-point Likert scale, ranging from 1 (strongly disagree) to 5 (strongly agree), with a score range of 8 to 40 to measure the perceived eHealth literacy of participants [19]. The validity of the Japanese version of eHEALS (J-eHEALS) has been determined, and a confirmatory factor analysis using data from a previous survey [2] was conducted. This analysis for the 8-item model suggested a good fit for the proposed model (goodness-of-fit index=0.988, confirmatory fit index=0.993, root mean square error of approximation=0.056), and the internal reliability of the test was confirmed using Cronbach alpha coefficient (Cronbach alpha=.93) [2].

### *Health Behavior*

Belloc and Breslow [38] have demonstrated the relationship between healthy behavior (including not smoking, regular physical exercise, moderate or no alcohol use, 7-8 hours' sleep, eating breakfast, and not eating between meals) and positive health status. Based on the work of Belloc and Breslow, Hagihara and Morimoto [39] added balanced nutrition defined as eating meals with balanced nutrition to their list of healthy behavior. Many studies have referred to the work of Breslow and Enstrom [40] and Breslow and Breslow [41] with regard to health behavior, and Morimoto and associates [42,43] for health status; therefore, we followed the studies of Belloc and Breslow [38] and Hagihara and Morimoto [39] to ensure the survey quality. In accordance with previous studies about health behavior and status, in this study we used a self-administered questionnaire, which included items related to cigarette smoking, physical exercise, alcohol consumption, sleeping hours, eating breakfast, eating between meals, and balanced nutrition, to assess health behavior [38,39].

With respect to smoking status, the questionnaire included an item about whether participants had ever smoked. Physical exercise was assessed by asking participants about their weekly frequency. Alcohol consumption was determined by inquiring about the weekly frequency. Number of sleeping hours was evaluated in terms of daily sleeping hours. Eating breakfast and eating between meals were categorized as follows: every day, almost every day, sometimes, and never. Balanced nutrition was grouped into three categories: eating a nutritionally balanced diet, eating with little regard to nutritional balance, and not eating a balanced diet. In accordance with previous studies [38-43], each health behavior was divided into one of two categories (good health behavior; poor health behavior) as follows: smoking cigarettes (never smoking; smoking) [38,40,41], regular physical exercise (twice or more a week; less than once a week) [39,42,43], moderate or no alcohol use (less than four times a week; five or more times a week) [38-43], sleeping hours (7-8 hour per night; ≤6 or ≥9 hour per night) [38-43], eating breakfast (almost every day or every day; sometimes or never) [38-43], not eating between meals (sometimes or never; almost every day or every day) [38,40,41], and balanced nutrition (eating a nutritionally balanced diet; eating with little regard to diet or not eating a balanced diet) [39,42,43].

## Statistical Analyses

We divided J-eHEALS score into one of two categories (high or low) relative to the median group value (median 24.02, IQR 19.19-27.82); we did so in accordance with previous studies that used eHEALS to analyze the association between eHealth literacy level and health behavior and health information seeking [30,33,44]. We employed a chi-square test to evaluate the proportional differences in each health behavior with respect to eHealth literacy. We conducted logistic regression analyses to estimate the association between each health behavior and eHealth literacy level. To determine the association of each health behavior with eHealth literacy level, we performed logistic regression analyses: we adjusted for sociodemographic variables (age group, marital status, educational attainment, and household income), frequency of Internet searching, and health behaviors that were statistically significant with respect to eHealth literacy in univariate analyses. We calculated adjusted odds ratios and 95% confidence intervals for each variable. In all analyses, $P<.05$ was considered statistically significant. We used PASW 19.0 to compute the statistics.



XSL•FO
RenderX

## Results

### Sociodemographic Variables and Frequency of Internet Searching

We received the data for 2142 adults (response rate: 21.04%, 2142/10,178) from the research company. We excluded respondents with incomplete data (missing rate: 1.26%, 27/2142) and therefore analyzed the data of 2115 adults who provided complete information for the study variables. Table 1 presents the characteristics of the respondents. In this study, the mean age of the participants was 39.7 years (SD 10.9); 49.74% (1052/2115) of the participants were male, 50.69% (1072/2115) had graduated from college or graduate school, and 23.74% (502/2115) were educated to a level below a high school diploma. Among the respondents, 47.28% (1000/2115) had a household income less than 5 million yen and 52.72% (1115/2115) earned 5 million yen or more, 58.06% (1228/2115) were married, and 72.06% (1524/2115) used the Internet to search for information every day. The mean J-eHEALS score was 23.4 (SD 6.4).

**Table 1.** Sociodemographic characteristics of participants (N=2115).

| Characteristics | n (%) |
|---|---|
| **Sex** | |
| Male | 1052 (49.74) |
| Female | 1063 (50.26) |
| **Age groups (years)** | |
| 20-29 | 527 (24.92) |
| 30-39 | 530 (25.06) |
| 40-49 | 531 (25.11) |
| 50-59 | 527 (24.92) |
| **Education level** | |
| ≤High school graduate | 502 (23.74) |
| Two-year college or career college | 541 (25.58) |
| ≥College graduate | 1072 (50.69) |
| **Household income (yen)** | |
| <5 million | 1000 (47.28) |
| ≥5 million | 1115 (52.72) |
| **Marital status** | |
| Not married | 887 (41.94) |
| Married | 1228 (58.06) |
| **Frequency of Internet searching (per week)** | |
| Every day | 1524 (72.06) |
| No every day | 591 (27.94) |

### Association Between eHealth Literacy and Health Behavior

In the univariate analyses, sleeping hours ($P$=.30), eating breakfast ($P$=.75), and eating snacks ($P$=.17) were not statistically significantly related to eHealth literacy level. However, cigarette smoking ($P$<.001), physical exercise ($P$=.001), alcohol consumption ($P$=.009), and balanced nutrition ($P$<.001) were significantly related to eHealth literacy level; those variables were included in the logistic regression model as controlling factors. Table 2 presents the results of the logistic regression analysis for the association between eHealth literacy and different types of health behavior. This table also shows the results of the logistic regression analysis for the association between eHealth literacy and each type of health behavior after controlling for covariates. After controlling for covariates, individuals with high eHealth literacy were significantly more likely to exhibit good health behavior with respect to physical exercise (adjusted odds ratio [AOR] 1.377, 95% CI 1.131-1.678) and eating a balanced diet (AOR 1.572, 95% CI 1.274-1.940) than people with low eHealth literacy. However, after controlling for covariates, we observed no significant association between eHEALS score and health behavior with respect to cigarette smoking, alcohol consumption, sleeping hours, eating breakfast, and eating between meals.



XSL•FO
RenderX

**Table 2.** Association between eHealth literacy and health behavior.

| Health behavior | OR (95% Cl) | *P* | AOR[a] (95% Cl) | *P* |
|---|---|---|---|---|
| Cigarette smoking | 0.866 (.729-1.029) | .10 | 0.862 (0.711-1.046) | .13 |
| Physical exercise | 1.470 (1.215-1.779) | <.001 | 1.377 (1.131-1.678) | .001 |
| Alcohol consumption | 0.847 (0.712-1.007) | .06 | 0.876 (0.727-1.055) | .88 |
| Sleeping hours | 1.039 (0.870-1.240) | .67 | 1.069 (0.890-1.282) | .48 |
| Eating breakfast | 1.198 (0.968-1.484) | .10 | 1.023 (0.814-1.286) | .84 |
| Eating between meals | 0.988 (0.829-1.177) | .89 | 1.044 (0.863-1.262) | .66 |
| Balanced nutrition | 1.764 (1.445-2.153) | <.001 | 1.572 (1.274-1.940) | <.001 |

[a] Adjusted for sociodemographic factors, frequency of Internet searching, and other health behaviors that were statistically significant with respect to eHealth literacy in univariate analyses.

# Discussion

## Principal Results

After controlling for sociodemographic variables, frequency of Internet searching, and other health behavior, this study found that adult Internet users with high eHealth literacy were significantly more likely to have good health behavior, such as physical exercise and balanced nutrition, than individuals with low eHealth literacy. However, we found no significant association between eHealth literacy and cigarette smoking, alcohol consumption, sleeping hours, eating breakfast, or eating between meals.

## Comparison With Previous Work

This study is the first to examine the association between eHealth literacy and the general health behaviors of cigarette smoking, physical exercise, alcohol consumption, sleeping habits, eating breakfast, eating between meals, and balanced nutrition among Internet adult users in Japan. After controlling for covariates, we found eHealth literacy to be associated with the good health behavior of physical exercise and balanced nutrition among Internet users. The results of this study support those of the IMeHU [28]. According to the IMeHU, individuals with higher eHealth literacy had greater motivation and efficacy in using the Internet for health information [28]. Previous investigations have shown that individuals with high eHealth literacy were more active consumers of online health information [2,30,33,34,45]—especially information related to exercise and nutrition [2,45]—than people with low eHealth literacy. This study reinforces the IMeHU findings, whereby eHealth literacy may mediate the association between social status and health behavior through the use of online health information [28].

This study demonstrates that high eHealth literacy may promote the healthy behavior of physical exercise and balanced nutrition among the general population of Internet users. One study among college students found that eHealth literacy promotes such healthy behavior as exercising; eating low-fat foods, low-sugar cereals, and vegetables and fruit; and always having quality sleep [36]. This study expands on those findings by focusing not on college students, but the general population. Because approximately 90% of general adults aged 30 to 59 years have used the Internet, it is becoming an effective intervention tool for promoting health behavior among ordinary people [2]. Therefore, to promote healthy behavior, including physical exercise and balanced nutrition, it is necessary to examine ways of enhancing eHealth literacy among adult Internet users.

One study has found that functional eHealth literacy and critical eHealth literacy displayed a positive predictive power with respect to eating and exercise behavior, although critical eHealth literacy was able only to positively predict sleep behaviors [36]. Hsu et al [36] found that functional eHealth literacy and interactive eHealth literacy were less influential with respect to health behavior than critical eHealth literacy as follows: according to involvement theory, critical eHealth literacy may motivate individuals more to seek and evaluate the quality of health information than functional eHealth literacy and interactive eHealth literacy. In this study, however, eHEALS was used as a single factor, and it did not include the three dimensions of functional, interactive, and critical eHealth literacy [2,19]. Thus, this study does not allow any discussion of the association of health behavior with functional, interactive, and critical eHealth literacy. Further research is needed to clarify the mechanisms whereby the three dimensions of eHealth literacy affect health behavior toward developing an effective eHealth literacy educational program for promoting healthy behavior among adult Internet users.

We found that the mean eHEALS score among Japanese Internet users was lower than that previously reported in the United States [34]. Our finding is in line with that of a previous study on health literacy [46]. In this study, the mean eHEALS scores was 23.4 (SD 6.4), which is similar to that in a previous investigation of Japanese Internet users [35]. Conversely, Tennant et al [34] determined the mean eHEALS score to be 29.05 (SD 5.75) among baby boomers and older adults in the United States (male: 54.8%; age: mean 67.46, SD 9.98 years). Although the participants in this study were older than the population in our study, the mean eHEALS scores they reported were higher. This difference may be explained by the argument of Nakayama et al [46], whereby the Japanese population have found it difficult to find health information on the Internet because there is no reliable, understandable, neutral, and comprehensive health website comparable to websites such as MedlinePlus (US National Library of Medicine).



XSL•FO
RenderX

### Limitations

This study has a number of limitations. First, the participants were recruited from a single Japanese Internet research service company; thus, the relationships assessed may have been biased because of the potentially nonrepresentative nature of this sample as general Japanese Internet users [47-49]. We made an equal allocation to the eight sample groups categorized by sex and age to minimize selection bias; however, there was still an unavoidable bias in the representativeness of the participants registered with the Internet research company. Among the registered participants, approximately 50% were male, approximately 55% were in their twenties and thirties, and approximately 45% had graduated from college or graduate school. By contrast, in the general Japanese population, one national survey found that approximately 30% of people were in their twenties and thirties among adults older than 20 years, and approximately 20% of people had graduated from college or graduate school [50]. Moreover, previous studies have indicated that respondents may have certain characteristics, such as having higher income, frequent access to the Internet, and being more likely to respond to a survey than the general Internet user population [48,49]. Therefore, it is necessary to note that the 2115 participants in this study were younger, more educated, had a higher income, and had greater Internet access than population of Internet users and the general population in Japan.

Second, health behavior and eHealth literacy were assessed only using a self-administered questionnaire. Inaccuracies in estimating health behavior and eHealth literacy level were thus unavoidable. Moreover, some studies have reported that eHEALS is inappropriate because it does not assess the ability to use Web 2.0 [18,21]. Therefore, it is necessary to improve the model of eHealth literacy to fit the rapid changes in the informational landscape created by Web 2.0 tools [18].

### Conclusions

Among Japanese adult Internet users, we found some health behaviors, including exercise and balanced nutrition, to be independently associated with eHealth literacy. In rapidly developing Internet user societies, further research is needed to identify the mechanisms linking eHealth literacy with health information seeking and health behavior toward designing effective strategies more precisely for promoting healthy behavior.

### Acknowledgments

This investigation was supported by Grants-in-Aid for Scientific Research (No: 22700688) from the Japan Society for the Promotion of Science and Global COE Program “Sport Sciences for the Promotion of Active Life” from the Japan Ministry of Education, Culture, Sports, Science and Technology.

### Conflicts of Interest

None declared.

### Multimedia Appendix 1

[PDF File (Adobe PDF File), 93KB - jmir_v18i7e192_app1.pdf ]

### References

1. Ministry of Public Management, Home Affairs, Posts and Telecommunications. Information & Communications Statistics Database. 2014. Communications usage trend survey in 2013 URL: http://www.soumu.go.jp/johotsusintokei/english/ [accessed 2016-06-26] [WebCite Cache ID 6iZEHQcpR]
2. Mitsutake S, Shibata A, Ishii K, Okazaki K, Oka K. [Developing Japanese version of the eHealth Literacy Scale (eHEALS)]. Nihon Koshu Eisei Zasshi 2011 May;58(5):361-371. [Medline: 21905612]
3. Fox S. Pew Research Center. 2013 Jan 15. Health online 2013: information triage URL: http://www.pewinternet.org/2013/01/15/information-triage/ [accessed 2016-06-26] [WebCite Cache ID 6iZEmYTWG]
4. Eysenbach G. The impact of the Internet on cancer outcomes. CA Cancer J Clin 2003;53(6):356-371. [Medline: 15224975]
5. Rice RE. Influences, usage, and outcomes of Internet health information searching: multivariate results from the Pew surveys. Int J Med Inform 2006 Jan;75(1):8-28. [doi: 10.1016/j.ijmedinf.2005.07.032] [Medline: 16125453]
6. Pingree S, Hawkins R, Baker T, duBenske L, Roberts LJ, Gustafson DH. The value of theory for enhancing and understanding e-health interventions. Am J Prev Med 2010 Jan;38(1):103-109 [FREE Full text] [doi: 10.1016/j.amepre.2009.09.035] [Medline: 20117565]
7. Ritterband LM, Thorndike FP, Cox DJ, Kovatchev BP, Gonder-Frederick LA. A behavior change model for internet interventions. Ann Behav Med 2009 Aug;38(1):18-27 [FREE Full text] [doi: 10.1007/s12160-009-9133-4] [Medline: 19802647]
8. Sørensen K, Van den Broucke S, Fullam J, Doyle G, Pelikan J, Slonska Z, (HLS-EU) Consortium Health Literacy Project European. Health literacy and public health: a systematic review and integration of definitions and models. BMC Public Health 2012;12:80 [FREE Full text] [doi: 10.1186/1471-2458-12-80] [Medline: 22276600]
9. Nielsen-Bohlman L, Panzer AM, Kindig DA. Health Literacy: A Prescription to End Confusion. Washington, DC: National Academies Press; 2004.



XSL•FO
RenderX

10. Berkman ND, Sheridan SL, Donahue KE, Halpern DJ, Crotty K. Low health literacy and health outcomes: an updated systematic review. Ann Intern Med 2011 Jul 19;155(2):97-107. [doi: 10.7326/0003-4819-155-2-201107190-00005] [Medline: 21768583]
11. The US Department of Health and Human Services. Healthy people 2020 URL: http://www.healthypeople.gov/2010/hp2020/Comments/default.asp [accessed 2016-06-26] [WebCite Cache ID 6iZFg8GPx]
12. Friedman DB, Hoffman-Goetz L, Arocha JF. Health literacy and the World Wide Web: comparing the readability of leading incident cancers on the Internet. Med Inform Internet Med 2006 Mar;31(1):67-87. [doi: 10.1080/14639230600628427] [Medline: 16754369]
13. Berland GK, Elliott MN, Morales LS, Algazy JI, Kravitz RL, Broder MS, et al. Health information on the Internet: accessibility, quality, and readability in English and Spanish. JAMA 2001;285(20):2612-2621 [FREE Full text] [Medline: 11368735]
14. Powers BJ, Trinh JV, Bosworth HB. Can this patient read and understand written health information? JAMA 2010 Jul 7;304(1):76-84. [doi: 10.1001/jama.2010.896] [Medline: 20606152]
15. Norman CD, Skinner HA. eHealth literacy: essential skills for consumer health in a networked world. J Med Internet Res 2006 Jun;8(2):e9 [FREE Full text] [doi: 10.2196/jmir.8.2.e9] [Medline: 16867972]
16. Chan CV, Kaufman DR. A framework for characterizing eHealth literacy demands and barriers. J Med Internet Res 2011 Nov;13(4):e94 [FREE Full text] [doi: 10.2196/jmir.1750] [Medline: 22094891]
17. Mitsutake S, Shibata A, Ishii K, Oka K. A review for the concept and recent trends of eHealth literacy. J Jpn Health Educ Promot 2012;20:221-232. [doi: 10.11260/kenkokyoiku.20.221]
18. Norman C. eHealth literacy 2.0: problems and opportunities with an evolving concept. J Med Internet Res 2011 Dec;13(4):e125 [FREE Full text] [doi: 10.2196/jmir.2035] [Medline: 22193243]
19. Norman CD, Skinner HA. eHEALS: The eHealth Literacy Scale. J Med Internet Res 2006 Nov;8(4):e27 [FREE Full text] [doi: 10.2196/jmir.8.4.e27] [Medline: 17213046]
20. van der Vaart R, van Deursen AJ, Drossaert CH, Taal E, van Dijk JA, van de Laar MA. Does the eHealth Literacy Scale (eHEALS) measure what it intends to measure? Validation of a Dutch version of the eHEALS in two adult populations. J Med Internet Res 2011;13(4):e86 [FREE Full text] [doi: 10.2196/jmir.1840] [Medline: 22071338]
21. van der Vaart R, Drossaert CH, de Heus M, Taal E, van de Laar MA. Measuring actual eHealth literacy among patients with rheumatic diseases: a qualitative analysis of problems encountered using Health 1.0 and Health 2.0 applications. J Med Internet Res 2013 Feb;15(2):e27 [FREE Full text] [doi: 10.2196/jmir.2428] [Medline: 23399720]
22. Chung S, Nahm E. Testing reliability and validity of the eHealth Literacy Scale (eHEALS) for older adults recruited online. Comput Inform Nurs 2015 Apr;33(4):150-156. [doi: 10.1097/CIN.0000000000000146] [Medline: 25783223]
23. Furstrand D, Kayser L. Development of the eHealth Literacy Assessment Toolkit, eHLA. Stud Health Technol Inform 2015;216:971. [Medline: 26262273]
24. Robinson C, Graham J. Perceived Internet health literacy of HIV-positive people through the provision of a computer and Internet health education intervention. Health Info Libr J 2010 Dec;27(4):295-303 [FREE Full text] [doi: 10.1111/j.1471-1842.2010.00898.x] [Medline: 21050372]
25. Xie B. Effects of an eHealth literacy intervention for older adults. J Med Internet Res 2011 Nov;13(4):e90 [FREE Full text] [doi: 10.2196/jmir.1880] [Medline: 22052161]
26. Paek H, Hove T. Social cognitive factors and perceived social influences that improve adolescent eHealth literacy. Health Commun 2012;27(8):727-737. [doi: 10.1080/10410236.2011.616627] [Medline: 22452551]
27. Watkins I, Xie B. eHealth literacy interventions for older adults: a systematic review of the literature. J Med Internet Res 2014;16(11):e225 [FREE Full text] [doi: 10.2196/jmir.3318] [Medline: 25386719]
28. Bodie GD, Dutta MJ. Understanding health literacy for strategic health marketing: eHealth literacy, health disparities, and the digital divide. Health Mark Q 2008 Jan;25(1-2):175-203. [doi: 10.1080/07359680802126301] [Medline: 18935884]
29. Cho J, Park D, Lee HE. Cognitive factors of using health apps: systematic analysis of relationships among health consciousness, health information orientation, eHealth literacy, and health app use efficacy. J Med Internet Res 2014;16(5):e125 [FREE Full text] [doi: 10.2196/jmir.3283] [Medline: 24824062]
30. Park H, Moon M, Baeg JH. Association of eHealth literacy with cancer information seeking and prior experience with cancer screening. Comput Inform Nurs 2014 Sep;32(9):458-463. [doi: 10.1097/CIN.0000000000000077] [Medline: 25105588]
31. Chang F, Chiu C, Chen P, Miao N, Lee C, Chiang J, et al. Relationship between parental and adolescent ehealth literacy and online health information seeking in Taiwan. Cyberpsychol Behav Soc Netw 2015 Oct;18(10):618-624. [doi: 10.1089/cyber.2015.0110] [Medline: 26375050]
32. Knapp C, Madden V, Marcu M, Wang H, Curtis C, Sloyer P, et al. Information seeking behaviors of parents whose children have life-threatening illnesses. Pediatr Blood Cancer 2011 May;56(5):805-811. [doi: 10.1002/pbc.22674] [Medline: 21370415]
33. Neter E, Brainin E. eHealth literacy: extending the digital divide to the realm of health information. J Med Internet Res 2012 Jan;14(1):e19 [FREE Full text] [doi: 10.2196/jmir.1619] [Medline: 22357448]




XSL•FO
RenderX

34. Tennant B, Stellefson M, Dodd V, Chaney B, Chaney D, Paige S, et al. eHealth literacy and Web 2.0 health information seeking behaviors among baby boomers and older adults. J Med Internet Res 2015;17(3):e70 [FREE Full text] [doi: 10.2196/jmir.3992] [Medline: 25783036]
35. Mitsutake S, Shibata A, Ishii K, Oka K. Association of eHealth literacy with colorectal cancer knowledge and screening practice among internet users in Japan. J Med Internet Res 2012;14(6):e153 [FREE Full text] [doi: 10.2196/jmir.1927] [Medline: 23149453]
36. Hsu W, Chiang C, Yang S. The effect of individual factors on health behaviors among college students: the mediating effects of eHealth literacy. J Med Internet Res 2014;16(12):e287 [FREE Full text] [doi: 10.2196/jmir.3542] [Medline: 25499086]
37. Ministry of Health, Labour and Welfare. 2012. Health Japan 21 (the second term) URL: http://www.mhlw.go.jp/seisakunitsuite/bunya/kenkou_iryou/kenkou/kenkounippon21/en/kenkounippon21/index.html [accessed 2016-06-26] [WebCite Cache ID 6iZFqCFlW]
38. Belloc NB, Breslow L. Relationship of physical health status and health practices. Prev Med 1972 Aug;1(3):409-421. [Medline: 5085007]
39. Hagihara A, Morimoto K. Personal health practices and attitudes toward nonsmokers' legal rights in Japan. Soc Sci Med 1991;33(6):717-721. [Medline: 1957192]
40. Breslow L, Enstrom JE. Persistence of health habits and their relationship to mortality. Prev Med 1980 Jul;9(4):469-483. [Medline: 7403016]
41. Breslow L, Breslow N. Health practices and disability: some evidence from Alameda County. Prev Med 1993 Jan;22(1):86-95. [doi: 10.1006/pmed.1993.1006] [Medline: 8475014]
42. Kusaka Y, Kondou H, Morimoto K. Healthy lifestyles are associated with higher natural killer cell activity. Prev Med 1992 Sep;21(5):602-615. [Medline: 1279663]
43. Suda M, Nakayama K, Morimoto K. Relationship between behavioral lifestyle and mental health status evaluated using the GHQ-28 and SDS questionnaires in Japanese factory workers. Ind Health 2007 Jun;45(3):467-473 [FREE Full text] [Medline: 17634696]
44. Park H, Lee E. Self-reported eHealth literacy among undergraduate nursing students in South Korea: a pilot study. Nurse Educ Today 2015 Feb;35(2):408-413. [doi: 10.1016/j.nedt.2014.10.022] [Medline: 25466791]
45. Beaudoin CE, Fernandez C, Wall JL, Farley TA. Promoting healthy eating and physical activity short-term effects of a mass media campaign. Am J Prev Med 2007 Mar;32(3):217-223. [doi: 10.1016/j.amepre.2006.11.002] [Medline: 17236742]
46. Nakayama K, Osaka W, Togari T, Ishikawa H, Yonekura Y, Sekido A, et al. Comprehensive health literacy in Japan is lower than in Europe: a validated Japanese-language assessment of health literacy. BMC Public Health 2015;15:505 [FREE Full text] [doi: 10.1186/s12889-015-1835-x] [Medline: 26001385]
47. Eysenbach G, Wyatt J. Using the Internet for surveys and health research. J Med Internet Res 2002 Nov;4(2):E13 [FREE Full text] [doi: 10.2196/jmir.4.2.e13] [Medline: 12554560]
48. Liao Y, Harada K, Shibata A, Ishii K, Oka K, Nakamura Y, et al. Perceived environmental factors associated with physical activity among normal-weight and overweight Japanese men. Int J Environ Res Public Health 2011 Apr;8(4):931-943 [FREE Full text] [doi: 10.3390/ijerph8040931] [Medline: 21695022]
49. Shibata A, Oka K, Harada K, Nakamura Y, Muraoka I. Psychological, social, and environmental factors to meeting physical activity recommendations among Japanese adults. Int J Behav Nutr Phys Act 2009;6:60 [FREE Full text] [doi: 10.1186/1479-5868-6-60] [Medline: 19715568]
50. Ministry of Internal Affairs and Communications. Statistics Japan. Population census 2010 URL: http://www.stat.go.jp/english/data/kokusei/ [accessed 2016-06-26] [WebCite Cache ID 6iZGA5GLX]


## Abbreviations

**AOR:** adjusted odds ratio
**eHEALS:** eHealth Literacy Scale
**IMeHU:** Integrative Model of eHealth Use


*Edited by G Eysenbach; submitted 08.12.15; peer-reviewed by J Cho, J Nguyen, J Jiang; comments to author 03.04.16; revised version received 16.05.16; accepted 04.06.16; published 18.07.16.*

*Please cite as:*
*Mitsutake S, Shibata A, Ishii K, Oka K*
*Associations of eHealth Literacy With Health Behavior Among Adult Internet Users*
*J Med Internet Res 2016;18(7):e192*
*URL: http://www.jmir.org/2016/7/e192/*
*doi:10.2196/jmir.5413*
*PMID:27432783*




XSL•FO
RenderX

©Seigo Mitsutake, Ai Shibata, Kaori ISHII, Koichiro Oka. Originally published in the Journal of Medical Internet Research (http://www.jmir.org), 18.07.2016. This is an open-access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/2.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on http://www.jmir.org/, as well as this copyright and license information must be included.



XSL•FO
RenderX

Original Paper

# Which Doctor to Trust: A Recommender System for Identifying the Right Doctors

Li Guo[1*], MS; Bo Jin[1,2*], PhD; Cuili Yao[2], MS; Haoyu Yang[2], BS; Degen Huang[1], PhD; Fei Wang[3], PhD

[1]School of Computer Science and Technology, Dalian University of Technology, Dalian, China

[2]School of Innovation and Entrepreneurship, Dalian University of Technology, Dalian, China

[3]Department of Healthcare Policy and Research, Weill Cornell Medical College, Cornell University, New York City, NY, United States

[*]these authors contributed equally

**Corresponding Author:**
Bo Jin, PhD
School of Innovation and Entrepreneurship
Dalian University of Technology
816 Yanjiao Building, Dalian University of Technolog
2 Linggong Road, High-Tech Zone
Dalian, 116024
China
Phone: 86 13084100305
Fax: 86 41184708599
Email: jinbo@dlut.edu.cn

## Abstract

**Background:** Key opinion leaders (KOLs) are people who can influence public opinion on a certain subject matter. In the field of medical and health informatics, it is critical to identify KOLs on various disease conditions. However, there have been very few studies on this topic.

**Objective:** We aimed to develop a recommender system for identifying KOLs for any specific disease with health care data mining.

**Methods:** We exploited an unsupervised aggregation approach for integrating various ranking features to identify doctors who have the potential to be KOLs on a range of diseases. We introduce the design, implementation, and deployment details of the recommender system. This system collects the professional footprints of doctors, such as papers in scientific journals, presentation activities, patient advocacy, and media exposure, and uses them as ranking features to identify KOLs.

**Results:** We collected the information of 2,381,750 doctors in China from 3,657,797 medical journal papers they published, together with their profiles, academic publications, and funding. The empirical results demonstrated that our system outperformed several benchmark systems by a significant margin. Moreover, we conducted a case study in a real-world system to verify the applicability of our proposed method.

**Conclusions:** Our results show that doctors' profiles and their academic publications are key data sources for identifying KOLs in the field of medical and health informatics. Moreover, we deployed the recommender system and applied the data service to a recommender system of the China-based Internet technology company NetEase. Patients can obtain authority ranking lists of doctors with this system on any given disease.

***(J Med Internet Res 2016;18(7):e186)*** doi:10.2196/jmir.6015

**KEYWORDS**

recommender systems; feature selection; rank aggregation; key opinion leaders

## Introduction

In the field of medical and health informatics, key opinion leaders (KOLs) are the doctors who can influence public opinion and lead the medical community through their research papers and clinic practices. These KOLs play important roles in the health care industry at every stage of their product life cycle. Therefore, there is a critical need for intelligent KOL identification services. Traditionally, consulting companies provided services for identifying KOLs by conducting user



XSL•FO
RenderX

surveys. These business solutions use only a limited number of information resources and focus on a small number of involved clients. Advances in informatics technologies have enabled us to collect large amounts of medical-related data [1], which in turn provide a new carrier for KOL identification. To this end, we conducted a large-scale quantitative analysis of multisource medical-related data and developed a recommender system for effectively identifying KOLs of any given type of disease by using such data.

KOL identification is also important to patients, since KOLs can influence which doctors patients want to approach. Several websites provide information on relevant doctors for patients, such as Yelp and Zocdoc. Yelp provides user reviews of doctors, but the quality of the reviews is not guaranteed. Zocdoc works primarily as a front end for managing a doctor's practice. The information used in both websites about doctors is relatively simple and not trustworthy.

In practice, one way to identify reliable KOLs is through referrals—in other words, the number of times a doctor is referred by another doctor. This can be treated as one type of social trust for doctors. In our method, we exploited coauthorship relationships and citation relationships to mimic such referrals. This process can be viewed as constructing doctor-centered networks from coauthorships and citations, which has been rarely studied (although there has been research on a patient-centered network [2]). On the other hand, although we cannot recognize good doctors only by counting their publications and all their citations [3], doctors whose papers are highly cited or who have published many papers in high-impact journals can promote their ideas and opinions to others more easily [4]. This is the same logic as that behind the PageRank algorithm for the Google search engine, which has also been used in the analysis of social network influence. In health informatics, KOL identification should encode objective and validated measurements of KOL activities, including academic publications, invited talks, quality of clinical research, patient evaluations, and media exposures. These activities should also be used as ranking features to identify KOLs.

The aim of this study was to develop a recommender system for identifying KOLs for any specific disease. Here we introduce the design, implementation, and deployment details of such a KOL identification system. Our system consists of 5 components: acquirement, integration, storage and access, modeling, and recommendation. The system is extensible and configurable, and has been deployed online for several months. In the recommendation component, we chose the profile of doctors, the expertise of doctors, and the social trust of doctors as the ranking features. The ranking function designed for KOL identification was constructed based on those features. We further developed an unsupervised ranking aggregation approach for KOL ranking. In a real-world deployment of our system, we also incorporated some external knowledge and optimized the settings of our system manually according to the recommendations of our operation team.

## Prior Work

KOLs are respected individuals who have a huge impact on other people's opinions, actions, and behaviors in a given social network [5]. Nowadays, people seek opinions and advice for supporting various decisions (eg, regarding medical treatment) from KOLs. Therefore, the key question is how to effectively and efficiently identify KOLs [6].

For academic research, there are mainly two categories of methods for identifying KOLs. The first category uses primary data, such as self-designation and peer identification [7]. The second uses secondary data, such as publications and social networks [8]. Primary data are more difficult to collect but are more accurate and effective [9]. There are also some combined methodologies using both primary and secondary data [5].

The number of business solutions encouraging KOL identification in the health care industry has also been increasing. For example, Thought Leader Select offers KOL identification, profiling, engagement planning, mapping, interviews, and surveys services to over two dozen of the world's largest biopharmaceutical and health care companies [10]. Moreover, a health care startup, HealthTap, constructed a doctor social graph to launch a service that maps doctors' connections [11]. Their graph, called DOConnect, has 25 million doctor referrals and was generated with big data technologies.

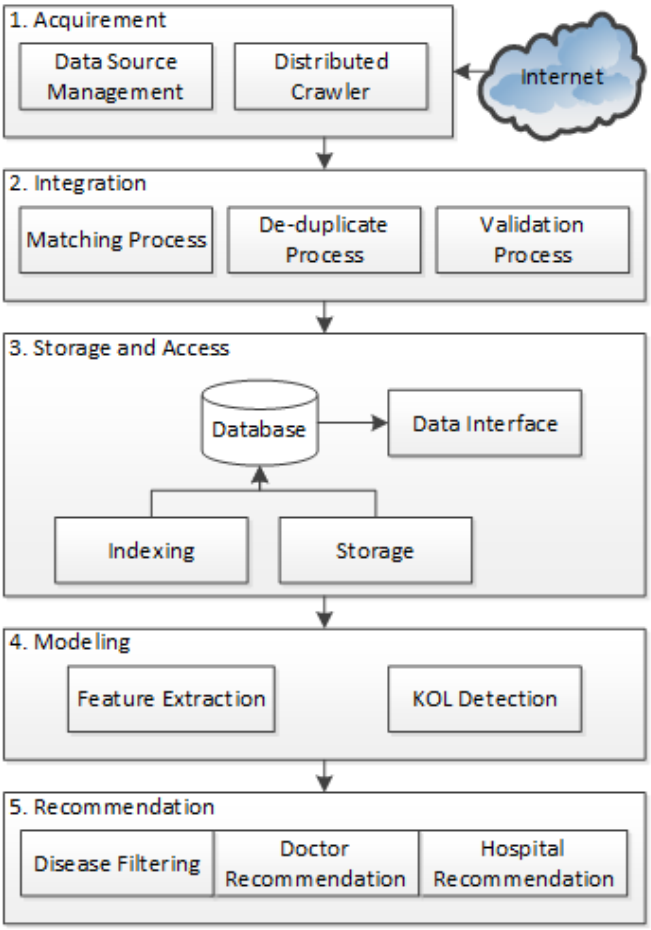
## System Overview

Figure 1 shows the architecture and workflow of our system, which consists of *acquirement*, *integration*, *storage and access*, *modeling* and *recommendation* stages.



XSL•FO
RenderX

**Figure 1.** Architectural overview of the key opinion leader (KOL) identification system.




### *Acquirement Stage*

This stage focuses on acquiring health care information from the Internet automatically. We developed an advanced Web crawler [12] for collecting the doctors' profiles and publications from multiple open data sources, which can be managed by rule-based operations.

### *Integration Stage*

This stage aims to integrate the doctors' profiles and publications through a data matching process. These data are further processed through a de-duplication and validation processes to improve their quality.

### *Storage and Access Stage*

This stage provides the capability of storing and indexing the integrated data. Specifically, we used MySQL for database storage and indexing, and provided a data access interface via Web service application programming interfaces.

### *KOL Identification Stage*

This stage identifies KOLs. In our system, this task is treated as a classic information retrieval task. Specifically, we used an unsupervised aggregation approach to integrate the ranking features of health care data for KOL identification.

### *Recommendation Stage*

This stage provides several recommendation services based on the results of KOL identification. Specifically, the system can return the ranked KOL list and corresponding hospital list as recommendations for users based on their personalized specifications, such as disease category. The recommendation results can be further filtered with the detailed disease names.

## Methods

### Design and Deployment

In this section, we discuss the design and deployment of our KOL identification system in detail. This system is based on a previously published study [13].

### *Data Acquisition*

To build our system, we used a Web crawler to collect large-scale health care-related data from multiple sources, including government public data, official hospital websites, professional health care websites, and medical companies' information systems.

A Web crawler is usually set in advance for a specific website design, and thus it is difficult to modify the crawler when the target site is changed. To meet the system requirement of multiple-source data acquisition, it is necessary to redesign the Web crawler. Here we present an advanced method for implementing Web crawler task management based on Jun [12], which Figure 2 shows. This method has the following steps: (1) initializing the link address of a webpage to be crawled by the client, (2) packaging the link address of the webpage to be crawled into a task request to the server by the client, (3) sending



XSL•FO
RenderX

an HTTP request from the server to the webpage to be crawled and returning the information required to the client, (4) receiving the information and processing the information on the client, (5) repeating the process and completing the webpage crawling in a crawling list sequentially. The proposed method provides a universal crawling framework for crawling different Internet content. In this way, crawlers for a special webpage can be quickly compiled, and thus the development can be much easier and more efficient. Furthermore, as the method is established based on the distributed Internet crawler framework, crawling efficiency can be further improved.

We also created a database to store the acquired data, which includes 54 tables (Figure 3). The structure of our database is extensible, and thus the database has the capability to incorporate more datasets in the future.

**Figure 2.** Data acquisition in key opinion leader identification system.

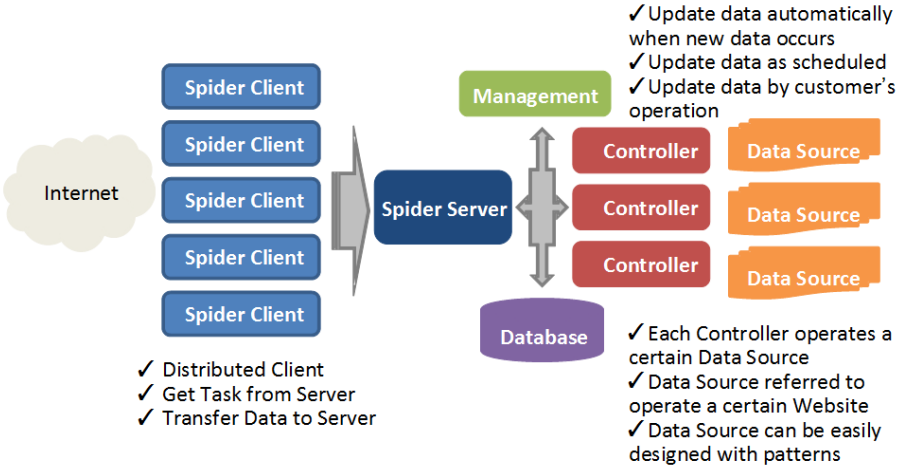
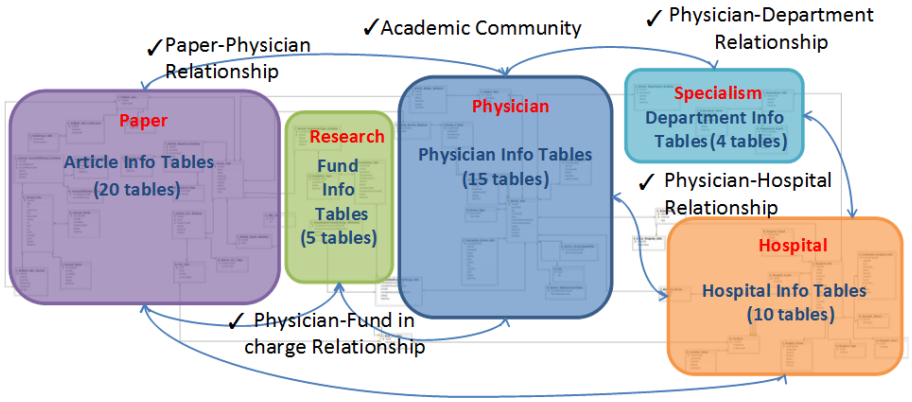



**Figure 3.** Database structure and information in the tables for data acquired by the key opinion leader identification system.




## *Data Processing*

As Figure 4 shows, our system processes data in the following 4 steps. The first step is to clean the acquired data. Since there is a lot of noise in the original data, we first identify the incomplete, incorrect, inaccurate, and irrelevant parts. Then, we clean, replace, modify, or delete such "dirty" data.

The second step is to match the multisource health care information. Since a hospital would have several names with different acronyms, the hospital names are matched using alias lists. Actually, the process of merging multisource information encounters a lot of name errors. The names of doctors are matched using Chinese pinyin (romanized Chinese ideograms), which can reduce written errors in Chinese characters.

The third step is to de-duplicate the doctors, since many names are duplicated. Therefore, we consider the same name appearing in the same hospital with the same specialty to be a single doctor, so that we can reduce the number of duplicated names.

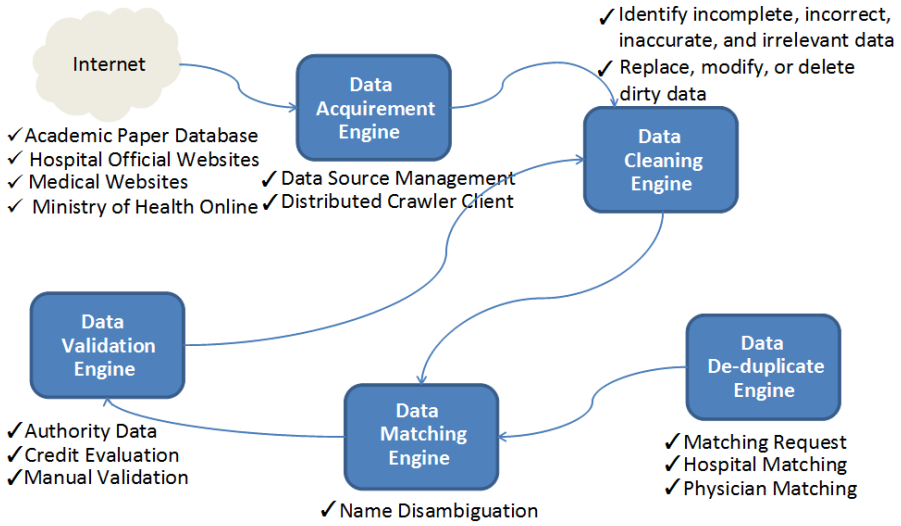
The final step is to validate the multisource doctor data. In particular, we validate the information's consistency across multiple sources. For any specific doctor, we retain her or his information from more reliable and more recent sources and discard the information from other sources when inconsistency appears. We also apply a manual check as the last step.

In our system, we use only academic papers in the domain of medicine to identify KOLs. Because not all authors of a paper are doctors, we match the paper's authors to the doctor dataset to identify the doctors more accurately.



XSL•FO
RenderX

**Figure 4.** Data preprocessing workflow in the key opinion leader identification system.




## *Data Analysis*

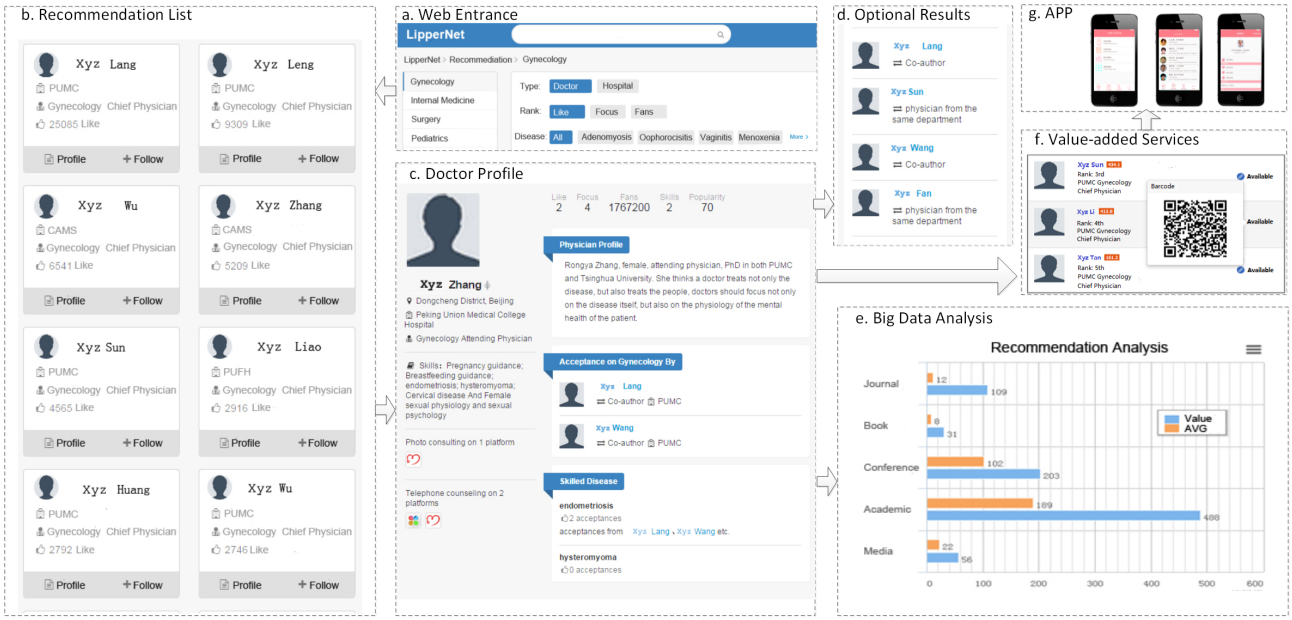
Our health care datasets contain almost all the registered doctors in China from the Chinese Ministry of Public Health. There are in total 2,381,750 doctors in the dataset. The profile of each doctor includes sex, age, specialty, title, employer, work experience, and resume. This information is collected from multiple sources. We have also crawled information for 106,021 hospitals in China. Hospitals are divided into 3 grades and 3 classes: grade III class A is the highest level, and grade I class C is the lowest level. Most doctors are employed in hospitals in grade II class A (41.5%) and grade III class A (31.7%).

In addition, our dataset contains information about all 1103 medical journals published in China. There are in total 3,657,797 papers (1980–2014) in the dataset. Information about each paper includes the journal name, publication date, volume, title, list of authors, authors' affiliations, classification identification, abstract, keywords, and references. Based on this information, we constructed a coauthorship network among doctors. For example, if 2 doctors coauthor at least one paper, then there will be a cooperative relationship between them. An analysis found that most doctors have no more than 50 coauthors, while the largest number of coauthors was over 300.

## *Web App*

Our system can produce recommendations for pharmaceutical companies and patients, and its Web-based front end enables content analysis and recommendations for users. Figure 5 shows screenshots from the Web app and the steps in making doctor recommendations.

**Figure 5.** Screenshots from the Web app showing doctor recommendation and content analysis functions.




## KOL Identification

In this section, we introduce the technical details of our KOL identification approach. First, we formally defined the problem of KOL identification in this study. Given a disease category $c$ as an element of the set $C$ and a set of doctors $D=\{ d_1, d_2, ..., d_n\}$, the problem of KOL identification is to find the top $K$



XSL•FO
RenderX

authoritative doctors in *D* for category *c*. Intuitively, this problem can be regarded as a classic information retrieval task, where the major challenge is how to define the ranking features for effectively linking doctors' expertise and disease categories. In the following we introduce the detailed ranking features used in our system and how to integrate these features for KOL identification.

### *Ranking Features for KOL Identification*

In our system, there are 3 types of ranking features for KOL identification, namely *doctor's profile*, *doctor's expertise*, and *social trust of the doctor*.

The doctor's profile is the basic descriptive information in his or her resume, such as demographic information, academic background, and professional activities. The system extracts 5 features based on the doctor profiles in our datasets: *professional duration*, *academic title* (eg, Full Professor), *professional title* (eg, Physician), and the *hospital level* where she or he works (eg, grade III class A). Table 1 (top) describes these features.

**Table 1.** Description of ranking features in the key opinion leader identification system.

| Feature type | Feature | Description |
|---|---|---|
| Profile features | Professional duration | Working years of the doctor |
| | Academic title | None, Assistant Professor, Associate Professor, Full Professor |
| | Professional title | None, Physician, Resident Physician, Physician in Charge, Associate Chief Physician, Chief Physician |
| | Hospital level | GI-A, GI-B, GI-C, GII-A, GII-B, GII-C, GIII-A, GIII-B, GIII-C[a] |
| Expertise features | Number of publications | Number of academic publications by a doctor in the given disease category |
| | Patient rating | Average rating of the doctor given by his or her patients |
| | Expertise label | Correspondence of the given disease category with a doctor's expertise labels |
| Social trust features | Coauthorship | Evaluation of the degree of collaboration between doctors |
| | Publication citation | Evaluation of the doctor's authority |
| | Social recognition | Evaluation of the degree of the doctor's social recognition |

[a]Grade and class of hospital (eg, grade I class A).

The doctor's expertise is used to evaluate the expertise level of a doctor with respect to the given disease category. Specifically, we extract 3 expertise features, described in the middle part of Table 1. The first feature is the *number of publications* a doctor has in a given disease category. The second feature is the doctor's average *patient rating* and can be used to evaluate his or her treatment in a given disease category. The third feature, *expertise label*, denotes the correspondence between a given disease category and a doctor's expertise labels.

To construct the expertise label feature, we label each doctor with a vector *y*. First, we select a group of doctors randomly and manually label each doctor with the disease category with which they are most experienced. Then we apply the label propagation algorithm [14] on multiple networks to predict labels corresponding to the expertise of all doctors in our datasets. After the labeling, we have $N_d$ label vectors in total, where $N_d$ denotes the number of doctors in the set of doctors *D*. Each label vector can be represented as an $N_c$-dimensional vector, where $N_c$ is the number of disease categories in *C*. Each dimension of the vector represents the extent to which a doctor is skilled in treating a specific disease category. If a doctor is perfect in treating a specific disease category, the corresponding value in the vector is set to 1; otherwise, if he or she is completely unable to treat the disease, the value is set to 0. Therefore, the expertise label score is computed as shown in equation (a) (Figure 6).

We evaluate the doctor's social trust with respect to a given disease category, which can be very useful for identifying KOLs among doctors. Specifically, we exploit 3 authority scores as social trust features in our system. The first score is *coauthorship*, which is defined to evaluate the degree of collaboration between doctors. Specifically, given a doctor *d* and all of his or her publications *P* in the given disease category *c*, the coauthorship is represented by the number of different authors in publication *P* except *d*. Generally, the more partners the doctor has, the stronger the academic influence she or he has. The second score is the *publication citation*, which is computed as the number of publications *P* that doctor *d* published in the given disease category *c* that were cited. The publication citation is a good performance indicator of his or her academic authority. Third, we extract *social recognition* as a feature to support the judgment of whether a doctor can be trusted. Specifically, social recognition is indicated by the number of the doctor's social fans. For example, the doctor's social recognition score $S_d$ is set to 2 if he or she has 20 social fans, the score is 3 for 100 social fans, and so on. However, not everyone has social networks, that is to say, not every doctor has social fans. If doctor *d* doesn't have a social network, then social recognition is set to 0. The feature descriptions are detailed at the bottom part of Table 1.



XSL•FO
RenderX

**Figure 6.** Equations used for the ranking functions.

(a) Expertise label score = $|\vec{y} \times \vec{I_c}|$

given a label vector $y$ of doctor $d$ and a disease vector $\vec{I_c}$ of a disease category $c$, where $\vec{I_c}$ is an $N_c$-dimensional zero vector, except the $c$-th value, which is set to 1

(b)

$$F(d,c) = \sum_{i=1} w_i \times f_i^{(d,c)}$$

where $F(d,c)$ is the ranking function, and weight $w_i \in [0,1]$ is the aggregation parameter of ranking features, which satisfies $\sum_{i=1} w_i = 1$,

(c) $\sigma_i(d,c) = (\pi_i(d,c) - \bar{\pi}(d,c))^2$

where $\sigma_i(d,c)$ is the variancelike measure, $\pi_i(d,c)$ is the ranking of doctor $d$ returned by $f_i^{(d,c)}$, and $\bar{\pi}(d,c)$ is the average ranking for doctor $d$

(d)

$$\arg\min_W \sum_{d \in D} \sum_{i=1} \sum_c w_i \cdot \sigma_i(d,c) + \lambda \|w\|^2,$$

$$\text{s.t.} \sum_{i=1} w_i = 1;\ \forall w_i \geq 0$$

where $\lambda$ is a regularization parameter to avoid overfitting during the optimization process for weight $w$

### *Ranking Function for KOL Identification*

After the above ranking features are constructed, the remaining task is how to integrate them for KOL identification. A common way is to define a linear ranking function with unknown feature weights as parameters, which are obtained from training data [15]. However, our data lack sufficient and reliable information that can be regarded as ground-truth ranking of doctors for each disease category, which makes it difficult to use a traditional supervised learning approach to obtain a ranking function. To solve this problem, in our system we use an unsupervised aggregation approach proposed by Zhu et al [16] for integrating ranking features.

Specifically, first we manually transform all categorical features into numerical values so that they can be used as scores for ranking doctors. For example, we transform the values of the feature *academic title* from none, Assistant Professor, Associate Professor, and Full Professor to 0, 1, 2, 3, and 4, respectively. Then, we implement normalization by subtracting the mean and dividing the standard deviation for all numerical features. After this, the ranking features of a given doctor-disease tuple ($d$, $c$) can be denoted as $\{f_1^{(d,c)}, f_2^{(d,c)}, ..., f_m^{(d,c)}\}$, where $m$ is the number of features we extracted. Meanwhile, the ranking function $F$ ($d$, $c$), which indicates the expertise score of $d$ in $c$, is defined by equation (b) (Figure 6). Given a set of doctors $D$, we select $n$ ranked lists with feature scores. Then $\pi_i(d, c)$ is the ranking of doctor $d$ returned by $f_i^{(d,c)}$, and $\pi$ ($d$, $c$) is the average ranking for doctor $d$. Thus, for feature $f_i^{(d,c)}$, consistency is calculated by the variance-like measure in equation (c) (Figure 6). The smaller $\sigma_i(d, c)$ is, the larger the weight, and vice versa, of $f_i^{(d,c)}$ should be assigned. Thus, the feature aggregation problem is defined as an optimization problem as shown by equation (d) (Figure 6).

The above problem can be solved by a gradient-based approach [16]. After learning the feature weights, we can rank the doctors with different disease categories for KOL recommendation. Our algorithm is based on the algorithm developed by Zhu et al [16] and Wang et al [17], which aims at minimizing the global inconsistency (reflected by the variance of ranking results) of all ranking measures.

## Results

In this section, we present the empirical results for validating the effectiveness of our system in terms of KOL identification with all of the data we crawled.

### Experimental Data

As mentioned above, there are many doctors in our system (2,381,750 doctors), but only a small percentage of the doctors can be identified as KOLs. Most doctors are at low-level health



XSL•FO
RenderX

organizations and we have little information for them. To evaluate our proposed method, we used a subset of our data as the experimental data, which we collected from We Doctor. This real-world data set includes 29,203 doctors in more than 7,000 expert teams all over China. Most of these doctors are experts in more than one discipline. Furthermore, each expert team has a leader, who can be treated as a KOL. That is to say, the leader of the expert team can influence at least the team members with his or her medicinal opinions.

First, we analyzed doctors' profiles and discovered that more than half of the doctors (up to 63.07%, 18,418/29,203) in the experimental dataset have senior titles, such as Chief Physician and Associate Chief Physician. In contrast, 35.73% (141,745/396,718) of doctors have senior titles in our full dataset from the top category of hospitals (grade III class A). This indicates that we used a subset of doctors who were more likely to be experts. Second, by analyzing patients' reviews, we found that most indicated the highest levels of satisfaction (ie, levels 8 and 9). A fairly large number of reviews reported dissatisfaction (ie, level 1). Few reviews indicated other levels of satisfaction. This indicates that patients tended to review doctors at the extremes, that is, either satisfied or dissatisfied, even for the experts. Third, an analysis of doctors' social media followers showed that most doctors had few followers, although some "star" doctors had a large number of followers.

## Evaluation of KOL Identification

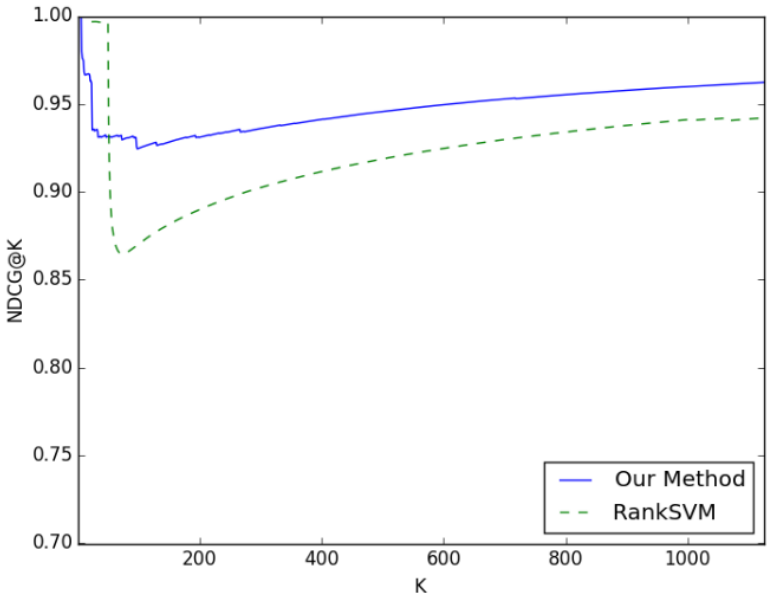
In China there is no public authority ranking list of doctors. Therefore, in this study, we evaluated the proposed approach with the doctor review and rating data gathered from our data service platform. We collected review logs of doctors and diseases entered into our data service between November 1, 2015 and January 31, 2016. There were 3496 review logs for 1133 doctors and 7823 review logs for 51 diseases.

We used RankSVM [18] as the baseline and used normalized discounted cumulative gain (NDCG) [19] to evaluate the performance of the recommendation result. NDCG indicates the ranking performance with a cutoff rank *K*. Figure 7 shows the recommendation performances of the two approaches. Our approach outperformed the baseline by a significant margin, especially for smaller *K* when *K* is larger than a threshold of 50.

We also did a focus group study with 1341 gynecologists in Beijing. To establish a reference standard, we invited 6 evaluators (3 faculty members with a medical background and 3 graduate students) to provide human judgments with scores of 4 (definite expertise), 3 (expertise), 2 (marginal expertise), 1 (little expertise), and 0 (no expertise). Group members based their judgments mainly on what they thought about the doctor's professional activities and reputation. After this user evaluation, each doctor was assigned a judgment score. We averaged the judgment scores and used them to rank the doctors. We selected the top 30 doctors to build the ground truth. Then we implemented our system and other systems (Haodaifu, Beijing, China; and DXY, Hangzhou, China) with similar functions in the evaluation dataset. We used the precision at 10 documents retrieved, R-precision, and mean average precision as performance measures [20].

Figure 8 shows the results of KOL identification. The evaluation terms (precision at 10 documents retrieved, R-precision, and mean average precision) of different diseases were averaged to obtain the experimental results. Our method performed better than the others.

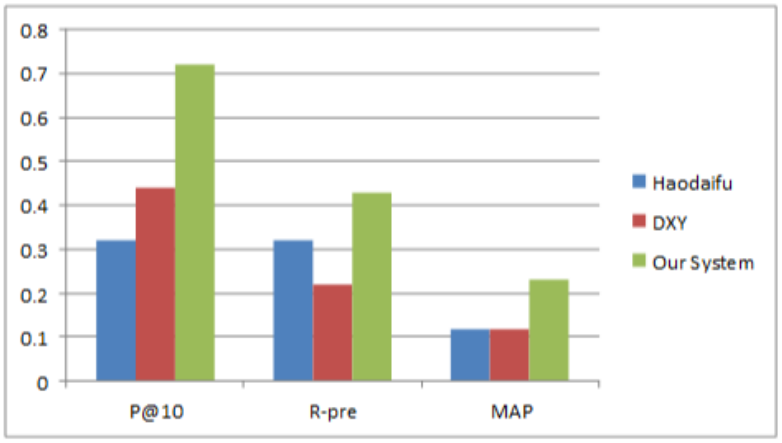
**Figure 7.** Evaluation by normalized discounted cumulative gain (NDCG) at cutoff rank K of recommendation performance by two approaches (RankSVM and the proposed method) based on data from November 1, 2015 to January 31, 2016.






XSL•FO
RenderX

**Figure 8.** Recommendation performance of different approaches on a small data subset evaluated by precision at 10 documents retrieved (P@10), R-precision (R-pre), and mean average precision (MAP).




## Discussion

We investigated and proposed new data mining models for KOL identification. Moreover, we have developed and deployed the KOL identification system. Over the past year, we have been deploying and testing our system online. The following section describes a case study that we applied to our system to verify the applicability of our proposed method.

### Case Study

We selected 5 diseases (adenomyosis, ovarian cyst, vaginitis, menoxenia, and cervicitis) from common gynecological categories for a case study. Table 2 shows the top 5 recommendation results of gynecologists in Beijing. There were 1341 gynecologists, most of whom were leading doctors for all of China. Our results show a high degree of overlap. Adenomyosis and menoxenia have the same doctor in the first position, as do ovarian cyst and vaginitis. This suggests that a leading doctor is ranked reasonably higher in similar or associated diseases, such as ovarian cyst and vaginitis. In contrast, the results of adenomyosis and ovarian cyst are quite different for the two diseases, which have less similarity or association. We also found that most of the recommended doctors were committee members of the gynecology branch of the Chinese Medical Association. For example, Jinghe Lang was the chairman of the gynecology branch. This validates our recommendation results.

**Table 2.** A case study of key opinion leader recommendations.

| Diseases | Doctors |
|---|---|
| Adenomyosis | Jinghe Lang, Jinhua Leng, Zhufeng Liu, Dawei Sun, Yingfang Zhou |
| Ovarian cyst | Zhaohui Liu, Fengzhi Feng, Bin Li, Jinsong Han |
| Vaginitis | Zhaohui Liu, Qinping Liao, Dai Zhang, Li Geng, Shuqing Jiang |
| Menoxenia | Jinghe Lang, Shan Deng, Ying Jin, Jian Shen, Ming Wu |
| Cervicitis | Qinping Liao, Li Geng, Lingying Wu, Wenhua Zhang |

We successfully applied our recommender system data service to NetEase, which is a leading China-based Internet technology company and is listed on NASDAQ as NTES.

### Conclusions

The KOL identification system we have developed can provide better KOL identification for pharmaceutical companies and patients. Our system integrates profiles of doctors and academic publications in the domain of medical science. This paper introduces the design, implementation, and deployment of our system. Specifically, we first acquired health care data from multiple sources using a Web crawler. Then we integrated the data into one system and preprocessed them using matching, de-duplication, and validation processes. We designed a storage system for the processed dataset and performed some basic statistical analyses on the dataset. Further, we proposed an approach of unsupervised ranking aggregation. Finally, this system can make recommendations to pharmaceutical companies and patients based on the proposed methods.

### Acknowledgments

The work was supported by the Natural Science Foundation of China (No. 91546123), the National Key R&D Plan by the Ministry of Science and Technology of China (No. 2016YFC0901900), and Fundamental Research Funds for the Central Universities (No. DUT15QY44).



XSL•FO
RenderX

## Conflicts of Interest

None declared.

## References


1. Xu H, Stenner SP, Doan S, Johnson KB, Waitman LR, Denny JC. MedEx: a medication information extraction system for clinical narratives. J Am Med Inform Assoc 2010;17(1):19-24 [FREE Full text] [doi: 10.1197/jamia.M3378] [Medline: 20064797]
2. Ohno-Machado L, Agha Z, Bell DS, Dahm L, Day ME, Doctor JN, et al. pSCANNER: patient-centered Scalable National Network for Effectiveness Research. J Am Med Inform Assoc 2014;21(4):621-626 [FREE Full text] [doi: 10.1136/amiajnl-2014-002751] [Medline: 24780722]
3. Werner R. The focus on bibliometrics makes papers less useful. Nature 2015 Jan 15;517(7534):245. [doi: 10.1038/517245a] [Medline: 25592498]
4. Uzuner Ö, Luo Y, Szolovits P. Evaluating the state-of-the-art in automatic de-identification. J Am Med Inform Assoc 2007;14(5):550-563 [FREE Full text] [doi: 10.1197/jamia.M2444] [Medline: 17600094]
5. Sharara H, Getoor L, Norton M. Active surveying: a probabilistic approach for identifying key opinion leaders. 2011 Presented at: The Twenty-Second international joint conference on Artificial Intelligence (IJCAI'11); July 16-22, 2011; Barcelona, Spain p. 1485-1490. [doi: 10.5591/978-1-57735-516-8/IJCAI11-250]
6. Kempe D, Kleinberg J, Tardos É. Maximizing the spread of influence through a social network. 2003 Presented at: The ninth ACM SIGKDD international conference on Knowledge discovery and data mining (KDD'03); August 24-27, 2003; Washington, DC, USA p. 137-146. [doi: 10.1145/956750.956769]
7. Weenen TC, Pronker ES, Commandeur HR, Claassen EH. Barriers to innovation in the medical nutrition industry: a quantitative key opinion leader analysis. PharmaNutrition 2013 Jul;1(3):79-85. [doi: 10.1016/j.phanu.2013.04.001]
8. Sun J, Tang J. A survey of models and algorithms for social influence analysis. In: Aggarwal CC, editor. Social Network Data Analytics. New York, NY: Springer; 2011:177-214.
9. Duan J, Zeng J, Luo B. Identification of opinion leaders based on user clustering and sentiment analysis. 2014 Presented at: The IEEE/WIC/ACM International Joint Conferences on Web Intelligence (WI) and Intelligent Agent Technologies (IAT); 2014; Warsaw, Poland p. 377-383. [doi: 10.1109/WI-IAT.2014.59]
10. Thought Leader Select. Chapel Hill, NC: Thought Leader Select; 2015. URL: http://www.thoughtleaderselect.com/
11. HealthTap. Palo Alto, CA: HealthTap; 2016. URL: https://www.healthtap.com/ [WebCite Cache ID 6inkU57dV]
12. Jin B. Method for Realizing Web Crawler Ttasks. CN Patent 102999549 A. 2013 Mar 27. URL: https://www.google.com/patents/CN102999549A?cl=en [accessed 2016-07-04] [WebCite Cache ID 6ikhKwnYs]
13. Jin B, Wang F, Ning X, Guo L, Zhong H, Xiong H. The current status of the health network in China: a real world case study with 106,021 hospitals. 2014 Presented at: Workshops at KDD 2014 Big CHat: Connected Health at Big Data Era; August 24-27, 2014; New York, NY, USA.
14. Kato T, Kashima H, Sugiyama M. Robust label propagation on multiple networks. IEEE Trans Neural Netw 2009 Jan;20(1):35-44. [doi: 10.1109/TNN.2008.2003354] [Medline: 19095533]
15. Zhou M, Chang KC. Unifying learning to rank and domain adaptation: enabling cross-task document scoring. 2014 Presented at: The 20th ACM SIGKDD international conference on Knowledge discovery and data mining; August 24-27, 2014; New York, NY, USA p. 781-790. [doi: 10.1145/2623330.2623739]
16. Zhu H, Xiong H, Ge Y, Chen E. Discovery of ranking fraud for mobile apps. IEEE Trans Knowl Data Eng 2015 Jan;27(1):74-87. [doi: 10.1109/TKDE.2014.2320733]
17. Wang F, Wang X, Li T. Generalized cluster aggregation. 2009 Presented at: The Twenty-first International Joint Conference on Artificial Intelligence (IJCAI-09); July 14-17, 2009; Pasadena, CA, USA p. 1279-1284.
18. Joachims T. Optimizing search engines using clickthrough data. 2002 Presented at: The eighth ACM SIGKDD international conference on Knowledge discovery and data mining (KDD'02); July 23-25, 2002; Edmonton, AB, Canada p. 133-142. [doi: 10.1145/775047.775067]
19. Manning CD, Raghavan P, Schütze H. Introduction to Information Retrieval. New York, NY: Cambridge University Press; 2008.
20. Buckley C, Voorhees EM. Retrieval evaluation with incomplete information. 2004 Presented at: The 27th annual international ACM SIGIR conference on Research and development in information retrieval; July 25-29, 2004; Sheffield, UK p. 25-32. [doi: 10.1145/1008992.1009000]


## Abbreviations

**KOL:** key opinion leader
**NDCG:** normalized discounted cumulative gain



XSL•FO
RenderX

*Edited by G Eysenbach; submitted 22.05.16; peer-reviewed by* 文 周*, P Gee; comments to author 10.06.16; revised version received 18.06.16; accepted 21.06.16; published 07.07.16.*

*Please cite as:*
*Guo L, Jin B, Yao C, Yang H, Huang D, Wang F*
*Which Doctor to Trust: A Recommender System for Identifying the Right Doctors*
*J Med Internet Res 2016;18(7):e186*
*URL: http://www.jmir.org/2016/7/e186/*
*doi:10.2196/jmir.6015*
*PMID:27390219*

©Li Guo, Bo Jin, Cuili Yao, Haoyu Yang, Degen Huang, Fei Wang. Originally published in the Journal of Medical Internet Research (http://www.jmir.org), 07.07.2016. This is an open-access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/2.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on http://www.jmir.org/, as well as this copyright and license information must be included.



XSL•FO
RenderX

Original Paper

# Influence of Intensified Supervision by Health Care Inspectorates on Online Patient Ratings of Hospitals: A Multilevel Study of More Than 43,000 Online Ratings

Rudolf Bertijn Kool[1], MD, PhD; Sophia Martine Kleefstra[2,3], PhD; Ine Borghans[2], PhD; Femke Atsma[1], PhD; Tom H van de Belt[4], PhD

[1]Radboud University Medical Center, Radboud Institute for Health Sciences, IQ Healthcare, Nijmegen, Netherlands

[2]Dutch Health Care Inspectorate (IGZ), Utrecht, Netherlands

[3]Academic Medical Center, Amsterdam, Netherlands

[4]Radboud University Medical Center, REshape Innovation Center, Nijmegen, Netherlands

**Corresponding Author:**
Rudolf Bertijn Kool, MD, PhD
Radboud University Medical Center
Radboud Institute for Health Sciences, IQ Healthcare
Postbus 9101, Huispost 114
Nijmegen,
Netherlands
Phone: 31 611531143
Fax: 31 243540166
Email: tijn.kool@radboudumc.nl

## Abstract

**Background:** In the Netherlands, hospitals with quality or safety issues are put under intensified supervision by the Dutch Health Care Inspectorate, which involves frequent announced and unannounced site visits and other measures. Patient rating sites are an upcoming phenomenon in health care. Patient reviews might be influenced by perceived quality including the media coverage of health care providers when the health care inspectorate imposes intensified supervision, but no data are available to show how these are related.

**Objective:** The aim of this study was to investigate whether and how being under intensified supervision of the health care inspectorate influences online patient ratings of hospitals.

**Methods:** We performed a longitudinal study using data from the patient rating site Zorgkaart Nederland, from January 1, 2010 to December 31, 2015. We compared data of 7 hospitals under intensified supervision with a control group of 28 hospitals. The dataset contained 43,856 ratings. We performed a multilevel logistic regression analysis to account for clustering of ratings within hospitals. Fixed effects in our analysis were hospital type, time, and the period of intensified supervision. Random effect was the hospital. The outcome variable was the dichotomized rating score.

**Results:** The period of intensified supervision was associated with a low rating score for the hospitals compared with control group hospitals; both 1 year before intensified supervision (odds ratio, OR, 1.67, 95% CI 1.06-2.63) and 1 year after (OR 1.79, 95% CI 1.14-2.81) the differences are significant. For all periods, the odds on a low rating score for hospitals under intensified supervision are higher than for the control group hospitals, corrected for time. Time is also associated with low rating scores, with decreasing ORs over time since 2010.

**Conclusions:** Hospitals that are confronted with intensified supervision by the health care inspectorate have lower ratings on patient rating sites. The scores are independent of the period: before, during, or just after the intervention by the health care inspectorate. Health care inspectorates might learn from these results because they indicate that the inspectorate identifies the same hospitals as “at risk” as the patients rate as underperformers.

***(J Med Internet Res 2016;18(7):e198)*** doi:10.2196/jmir.5884

**KEYWORDS**

rating sites; supervision; social media; online reviews; hospitals



XSL•FO
RenderX

## Introduction

Patient rating sites (PRSs) are an upcoming phenomenon in health care [1]. In many countries, websites such as RateMDs, Vitals, and Zocdoc in the United States, Jameda in Germany, and NHS Choices in the United Kingdom have created a platform for patients to share their experiences with health care providers. The number of ratings is growing and it comprehends all kinds of care. Nevertheless, the usefulness of PRSs for health care is being discussed [2-5]. Recently, a scoping review showed a growing body of literature on positive relationships between ratings on PRSs and indicators of quality of care such as patient satisfaction, mortality, and readmissions [6,7]. The content of PRSs is used for several purposes including supervision by the Dutch Health Care Inspectorate (Inspectie voor de Gezondheidszorg or IGZ) since 2015 [8].

Patient reviews are influenced not only by the type of care received by the patient and the way health care is delivered, but probably also by external stimuli such as media coverage. Research has shown that patients, especially the local ones, are influenced in their trust in health care providers by media attention [9]. In the Netherlands, media generally pay close attention to health care providers where "something might be wrong." Recent examples are 2 hospitals that came under intensified supervision of the IGZ after it noted patient safety problems [10,11]. Intensified supervision means frequent announced and unannounced site visits and consultation with the board, and it will be ended when structural improvements of the quality and safety of care have been proven and the board shows to be in control, see Textbox 1 [12]. In particular, health care magazines and local newspapers and sometimes local or even national radio and television report the intensified supervision and highlight it with special coverage. This could influence the opinion of patients of the health care providers and so their ratings on PRSs.

The fact that a health care inspectorate uses patient reviews as one of the components in risk detection, and that at the same time patient reviews could be influenced by intervention by the health care inspectorate, raises the question of how patient ratings of health care providers change after publication of intensified supervision by the inspectorate. There is no existing research on the extent to which patient ratings of health care providers are affected by intensified supervision of a health care inspectorate. One might expect that mean patient ratings of health care providers under intensified supervision are generally low compared with other health care providers before intensified supervision, because of the quality problems that are found. Patients and their relatives may use low ratings to draw attention to the problems they have experienced. Our first hypothesis was therefore that health care providers under intensified supervision would have a relative low mean overall patient rating in the period before intensified supervision compared with comparable providers.

Second, the mean rating of health care providers under intensified supervision might probably increase during the intensified supervision compared with the period before intensified supervision. Although negative media attention can ruin a reputation, previous research has also shown that the trust of patients in their doctors is high [13] and trust influences loyalty [14]. Patients might support their health care provider by sharing positive experiences and higher ratings. Third, we assumed that after the intensified supervision, when most of the publicity is gone, mean ratings of health care providers would be stable and comparable with others.

To investigate the aforementioned assumptions, we performed a multilevel study with the following research question: To which extent is intensified supervision associated with online hospital patient ratings? Because of the limited number of ratings of other health care providers, we focused on only hospitals in this study.

**Textbox 1.** The Dutch Health Care Inspectorate.

The Dutch Health Care Inspectorate (Inspectie voor de Gezondheidszorg or IGZ) is an agency under the Ministry of Health, Welfare and Sport. It is the official regulatory body charged with supervising the quality and safety of health care services, prevention activities, and medical products in the Netherlands. The IGZ has organized its supervision in several ways to ensure compliance with (professional) standards and guidelines and to ensure patient safety. The two most important methods are incident-based supervision and analyses of various types of risk information, also known as risk-based supervision.

The IGZ can also impose intensified supervision on a provider of care, entirely or one of its departments, if the reports from the inspectors, any reports and analyses of calamities, and/or the risk indicators show high risks for quality and/or safety of care and when there is insufficient faith in the strength and effectiveness of the board to realize improvements on time. Intensified supervision includes frequent announced and unannounced site visits and consultation with the board. Intensified supervision will be ended when structural improvements of the quality and safety of care have been proven and the board shows to be in control. When deciding upon the most appropriate enforcement measure, the inspectorate will take the following variables into account:

- the 5 D's: dissatisfaction, discomfort, disease, disability, and death (internationally recognized criteria);
- the number of people at risk (ie, a large, medium, or small risk group);
- the manner in which care provision is organized and structured with a view to quality and safety outcomes (poor, moderate, good); and
- the attitude of the care provider (ignorance, incompetence, noncompliance).



XSL•FO
RenderX

## Methods

### Study Design

We performed an observational study using publicly available data.

### Data Sources

First, we scanned the website of the IGZ where it publicly announces the providers that will be monitored by intensified supervision in order to arrange a list of all intensified supervision hospitals. We collected the names of these hospitals from January 2010 to December 2015, including the exact start and end date of the intensified supervision. In this period the IGZ decided to monitor 7 entire hospitals intensively by intensified supervision and to end it. We excluded 2 hospitals in which only 1 department was monitored by the inspectorate.

Second, we used data from the biggest PRS in the Netherlands, Zorgkaart Nederland, with more than 272,000 ratings and more than a million visitors per month in 2015. We used the publicly available data from January 1, 2010 to December 31, 2015, in which period the IGZ did not yet use hospital ratings to decide on intensified supervision. On Zorgkaart Nederland, patients can rate health care providers on 6 aspects, namely, accommodation, appointments, treatment, information, listening, and professionals, based on a scale from 1-10, where "1" stands for "extremely poor" and "10" for "extremely good." An overall score is automatically calculated (rounded to the nearest 0.1 percentage point). We used the number of ratings, the percentage of reviews with a score lower than 6.5, and the mean rating score on Zorgkaart Nederland given by patients. The 6.5 threshold is based on the theory of the Net Promoter Score [15]. This theory considers the scores 9 and 10 as positive "promoters," the scores 7 and 8 as neutral, passively satisfied, and the scores 0 to 6 ratings as "detractors," or negative recommendations. Because we were primarily interested in the "detractors," as these ratings might tell something about patient safety, we dichotomized the dependent variable rating scores: scores higher than and equal to 6.5 were labeled "0" and scores lower than 6.5 were labeled "1."

### Analysis

We compared the data of the 7 intensified supervision hospitals with a control group of 28 hospitals. In the Netherlands, 3 types of hospitals exist: relatively small general acute care hospitals, the bigger teaching hospitals, and the major academic hospitals. The control group hospitals were purposively sampled in the same region (North, South, East, and West) and same type as the intensified supervision hospitals and also on having the most ratings, thus guaranteeing a sufficient number of ratings. The total number of hospitals included in the dataset was 35 with 43,856 ratings. We analyzed the data for the different categories of hospitals because we could expect a difference in rating scores; patient satisfaction does differ between small and major hospitals [16]. We also analyzed the data for every year in order to visualize time effects.

Because we expected ratings to be influenced quite a while before the intensified supervision, we categorized the intensified supervision period into 5 categories: the period before 1 year before intensified supervision, the 12 months before intensified supervision, during intensified supervision, the 12 months after intensified supervision, and the period after 1 year after intensified supervision. The period during intensified supervision varied according to the decision of the inspectorate to prolong intensified supervision (3-12 months; mean 7 months).

We performed a longitudinal logistic regression analysis (mixed model) to analyze whether periods of intensified supervision lead to lower patient ratings. To account for clustering of ratings within hospitals, "hospital" was included in the model as the random effect. The outcome variable was the dichotomized rating score.

Determinants in our analysis were the period of intensified supervision, hospital type, and the course of ratings over time (variable "time" in years). We included the factor time to analyze time trends that occur anyway, irrespective of intensified supervision.

### Ethical Approval

No ethical approval was needed because we used publicly available data and no persons were directly involved.

## Results

Table 1 lists the numbers of ratings for intensified supervision hospitals and control group hospitals per period and per hospital type. Table 2 presents the mean rating score and percentage of ratings lower than 6.5 per period for intensified supervision hospitals and control group hospitals. Table 3 presents the mean rating in time and the percentage of ratings lower that 6.5, for both groups.



XSL•FO
RenderX

**Table 1.** Numbers of rating scores per hospital type and period.

| Hospitals and period | General acute care hospital | | Teaching hospital | | Academic hospital | | Total | |
|---|---|---|---|---|---|---|---|---|
| | N hospitals | N ratings | N hospitals | N ratings | N hospitals | N ratings | N hospitals | N ratings |
| Control group hospitals | 16 | 17,569 | 8 | 17,926 | 4 | 3967 | 28 | 39,462 |
| Intensified supervision hospitals | 4 | | 2 | | 1 | | 7 | |
| Period before 1 year before intensified supervision | | 499 | | 161 | | 81 | | 741 |
| 1 year before intensified supervision | | 303 | | 165 | | 82 | | 550 |
| During intensified supervision | | 285 | | 106 | | 53 | | 444 |
| 1 year after intensified supervision | | 272 | | 313 | | 94 | | 679 |
| Period after 1 year after intensified supervision | | 604 | | 1144 | | 232 | | 1980 |
| Total | 20 | 19,532 | 10 | 19,815 | 5 | 4509 | 35 | 43,856 |

**Table 2.** Mean rating score and percentage of ratings <6.5 per period for both intensified supervision and control group hospitals.

| Hospitals and period | Mean rating score | Percentage of ratings <6.5 |
|---|---|---|
| Control group hospitals | 8.5 | 8.3% |
| Intensified supervision hospitals | | |
| Period before 1 year before intensified supervision | 7.9 | 18.4% |
| 1 year before intensified supervision | 8.1 | 16.5% |
| During intensified supervision | 8.2 | 14.9% |
| 1 year after intensified supervision | 8.2 | 15.9% |
| Period after 1 year after intensified supervision | 8.5 | 10.6% |

**Table 3.** Mean rating in time and percentage <6.5 for both intensified supervision and control group hospitals.

| Year | Mean rating (%<6.5) intensified supervision hospitals | Mean rating (%<6.5) control group hospitals |
|---|---|---|
| 2010 | 7.7 (19.1) | 7.6 (19.9) |
| 2011 | 7.8 (22.6) | 8.2 (13.6) |
| 2012 | 8.2 (15.6) | 8.5 (8.6) |
| 2013 | 8.5 (10.7) | 8.7 (6.4) |
| 2014 | 8.4 (12.6) | 8.6 (7.9) |
| 2015 | 8.4 (12.8) | 8.7 (7.2) |

Table 4 presents the results of the multilevel analysis. The period of supervision is associated with a low rating score for intensified supervision hospitals compared with control group hospitals: both 1 year before intensified supervision (odds ratio, OR, 1.67, 95% CI 1.06-2.63) and 1 year after intensified supervision (OR 1.79, 95% CI 1.14-2.81) differ significantly. For all periods the odds on a low rating score for intensified supervision hospitals are higher than for the control group hospitals, corrected for time. The proportion of low rating scores decreased over time since 2010.



XSL•FO
RenderX

**Table 4.** Effect of intensified supervision period on low patient ratings, adjusted for time and type of hospital.

| Variables | Categories | Odds ratios | 95% confidence interval | |
|---|---|---|---|---|
| | | | Lower | Upper |
| | Control group hospitals | 1.0[a] | | |
| Intensified supervision period | Period before 1 year before intensified supervision | 1.29 | 0.825 | 2.028 |
| | 1 year before intensified supervision | 1.67 | 1.059 | 2.634 |
| | During intensified supervision | 1.55 | 0.964 | 2.506 |
| | 1 year after intensified supervision | 1.79 | 1.143 | 2.805 |
| | Period after 1 year after intensified supervision | 1.41 | 0.921 | 2.161 |
| Time | 2010 | 1.0[a] | | |
| | 2011 | 0.77 | 0.656 | 0.908 |
| | 2012 | 0.49 | 0.418 | 0.572 |
| | 2013 | 0.35 | 0.3 | 0.412 |
| | 2014 | 0.45 | 0.387 | 0.522 |
| | 2015 | 0.42 | 0.364 | 0.492 |
| Hospital type | General acute | 0.86 | 0.597 | 1.232 |
| | Academic | 1.1 | 0.659 | 1.836 |
| | Teaching | 1.0[a] | | |

[a] Reference category.

## Discussion

On the basis of the results of this study, we can confirm the hypothesis that the average rating of intensified supervision hospitals before the intensified supervision started is lower compared with the control group. We found also that patient ratings were low not only before but also during and just after the intensified supervision: the scores are continuously relatively low. It seems that only during the period after a year after intensified supervision the ratings are comparable with the control group hospitals. This result might assure the inspectorate that intensified supervision does not influence the patient ratings that it uses for supervision significantly in the short term. The results do not indicate that the mean rating will increase during the intensified supervision, for example, because of the loyalty of patients.

Health care inspectorates might learn from these results because they indicate that the inspectorate identifies the same hospitals as “at risk” as the patients rate as underperformers. This can be seen as another indication of the opportunities for patients to identify patient safety problems [17,18]. Monthly monitoring of scores on PRSs by health care inspectorates or other quality monitoring organizations could be of additional value in identifying health care providers at risk. This is also in line with the results of several studies that show correlations between patient rating scores and quality indicators, although the correlations were mostly weak and sometimes inconsistent [6]. The potential contribution of patient rating scores to health care governance supports the initiatives of health care inspectorates already using these online scores in their daily supervision [8].

Although not part of the research question, we identified a trend in patient rating scores. Compared with the start in 2010, the mean overall ratings on PRS Zorgkaart Nederland have increased significantly in 2015. To the best of our knowledge, there has not been an analysis of the rating trend over several years. This might be due to relative low ratings of pioneers at the start of the PRS. Now that the PRS is used more, it might attract a broader public with more positive ratings in general. In total, 92% of the ratings in our dataset are positive (≥6.5), which is comparable with approximately 90% of ratings found in other research [19].

### Strengths and Limitations

An important strength of this study is the major database used with 43,856 online ratings of 35 hospitals, more than a third of all Dutch hospitals.

In addition, it takes into account the opportunities of using patient experiences in supervision of health care. Especially with the increasing use of rating sites in the near future when the generation socialized with social media (eg, Facebook and Twitter) reaches the age in which health questions and doctors become significant, these kinds of sources might become even more relevant for patients, physicians, and other stakeholders such as health care inspectorates [2,4,20]. An increasing number of patients are given a voice to their experiences in order to identify patient safety risks. Health care inspectorates all over the world might benefit from this and better involve citizens in health care governance.

A limitation of the study is that we only selected the hospitals with sufficient reviews in the control group. We cannot rule out that this was a selected group of hospitals. Furthermore, the



XSL•FO
RenderX

number of hospitals with intensified supervision was low. There were only 7 hospitals confronted with intensified supervision. However, we analyzed the data on the level of patient ratings, which provided us with a dataset with enough power (n=43,856). The number of hospitals under supervision is only a small part of all providers in the Dutch health care with intensified supervision. In 2013 and 2014 it was concerned with 47 health care organizations of which 3 were hospitals. A second limitation is that at the moment the number of ratings of the other health care providers, mainly long-term elderly care, is too low to use for analysis. However, this could change rapidly. The branch organization of long-term elderly care announced in July 2015 the decision to cooperate with the Dutch Patient and Consumer Federation, the owner of Zorgkaart Nederland, in order to increase the number of reviews substantially by collecting ratings via interviews.

### Future Research

We expect that an increasing number of patients will share their experiences on PRSs and a growing number of patients will use those experiences in their judgment and choice of health care providers. This study suggests that all those experiences could be useful in estimating the quality of care because of the interesting association with the judgment of health care inspectors. However, it takes extensive research to understand this relationship better. Because of this association, this study might encourage health care inspectorates in experimenting with civilians as layman inspectors. It would also be interesting to know what underlying aspect of the IGZ's decision the negative patient reviews may be correlating with.

Exploring the reasons for low patient ratings in general will be necessary to give hospitals insight into how to improve their ratings. This could be investigated by closely studying the texts of the reviews, for example, by using Web-based text processing tools [21]. It might also be instructive for hospitals and IGZ to explore if any of the 6 aspects that go into the overall patient rating on Zorgkaart Nederland are more specifically correlated to intensified supervision rather than the overall score.

It might also be useful to explore the use of online patient ratings by health care parties other than inspectorates, for example, health insurance companies. They could start using rating scores to select preferred providers in their purchase of care.

Finally, research on the influence of supervisory activities should also be performed with other health care providers than hospitals, such as long-term care institutions, under the condition of sufficient number of reviews. Moreover, it would be relevant to repeat the study in other countries to investigate whether a different system of supervision or the presence of several PRSs influences the results.

### Conclusions

Hospitals that are confronted with intensified supervision by the health care inspectorate have lower ratings on PRSs. Health care inspectorates might learn from these results because they indicate that the inspectorate identifies the same hospitals as “at risk” as the patients rate as underperformers. More research with more ratings also in other parts of health care and other countries is needed to explore further the association between ratings on PRSs and the quality judgment of a health care inspectorate.

### Acknowledgments

We would like to thank the Dutch Patient and Consumer Federation (NPCF) for making available the data from Zorgkaart Nederland. Funding for this project was received from NWO/ZonMw.

### Authors' Contributions

RBK designed the study and drafted the manuscript. SMK and FA analyzed and interpreted the data. All authors reviewed and revised the manuscript at various stages to its final version and have read and approved the final manuscript.

### Conflicts of Interest

None declared.

### References

1. Burkle CM, Keegan MT. Popularity of internet physician rating sites and their apparent influence on patients’ choices of physicians. BMC Health Services Research 2015 Sep 26;15:416. [doi: 10.1186/s12913-015-1099-2]
2. Greaves F, Ramirez-Cano D, Millett C, Darzi A, Donaldson L. Harnessing the cloud of patient experience: using social media to detect poor quality healthcare. BMJ Qual Saf 2013 Mar;22(3):251-255. [doi: 10.1136/bmjqs-2012-001527] [Medline: 23349387]
3. Galizzi MM, Miraldo M, Stavropoulou C, Desai M, Jayatunga W, Joshi M, et al. Who is more likely to use doctor-rating websites, and why? A cross-sectional study in London. BMJ Open 2012;2(6):e001493. [doi: 10.1136/bmjopen-2012-001493]
4. Reimann S, Strech D. The representation of patient experience and satisfaction in physician rating sites. A criteria-based analysis of English- and German-language sites. BMC Health Serv Res 2010 Dec 7;10:332 [FREE Full text] [doi: 10.1186/1472-6963-10-332] [Medline: 21138579]
5. Strech D. Ethical principles for physician rating sites. J Med Internet Res 2011;13(4):e113 [FREE Full text] [doi: 10.2196/jmir.1899] [Medline: 22146737]



XSL•FO
RenderX

6. Verhoef LM, Van de Belt TH, Engelen LJ, Schoonhoven L, Kool RB. Social media and rating sites as tools to understanding quality of care: a scoping review. J Med Internet Res 2014;16(2):e56 [FREE Full text] [doi: 10.2196/jmir.3024] [Medline: 24566844]
7. Emmert M, Adelhardt T, Sander U, Wambach V, Lindenthal J. A cross-sectional study assessing the association between online ratings and structural and quality of care measures: results from two German physician rating websites. BMC Health Services Research 2015 Sep 24;15:414.
8. van de Belt TH, Engelen LJ, Verhoef LM, van der Weide MJ, Schoonhoven L, Kool RB. Using patient experiences on Dutch social media to supervise health care services: exploratory study. J Med Internet Res 2015 Jan;17(1):e7 [FREE Full text] [doi: 10.2196/jmir.3906] [Medline: 25592481]
9. van der Schee E, de Jong JD, Groenewegen PP. The influence of a local, media covered hospital incident on public trust in health care. Eur J Public Health 2012 Aug;22(4):459-464 [FREE Full text] [doi: 10.1093/eurpub/ckr033] [Medline: 21441560]
10. Sheldon T. Dutch microbiologists are disciplined for delays during pneumonia outbreak. BMJ 2012 Jan 30;344:e755.
11. Sheldon T. Senior managers resign over conflict with doctors. BMJ 2012 Aug 31;345:e5895.
12. Hout FA, Nienhuis ED, Robben PB, Frederiks BJ, Legemaate J. Supervision by the Dutch Healthcare Inspectorate. Eur J Health Law 2010 Sep;17(4):347-360. [Medline: 20873515]
13. Van der Schee E, Groenewegen PP, Friele RD. Public trust in health care: a performance indicator? J Health Organ Manag 2006;20(5):468-476. [doi: 10.1108/14777260610701821] [Medline: 17087406]
14. Torres E, Vasquez-Parraga AZ, Barra C. The path of patient loyalty and the role of doctor reputation. Health Mark Q 2009;26(3):183-197. [doi: 10.1080/07359680903263565] [Medline: 19813122]
15. Reichheld FF. The one number you need to grow. Harv Bus Rev 2003 Dec;81(12):46-54, 124. [Medline: 14712543]
16. Hekkert KD, Cihangir S, Kleefstra SM, van den Berg B, Kool RB. Patient satisfaction revisited: a multilevel approach. Soc Sci Med 2009 Jul;69(1):68-75. [doi: 10.1016/j.socscimed.2009.04.016] [Medline: 19446942]
17. Daniels JP, Hunc K, Cochrane DD, Carr R, Shaw NT, Taylor A, et al. Identification by families of pediatric adverse events and near misses overlooked by health care providers. CMAJ 2012 Jan 10;184(1):29-34 [FREE Full text] [doi: 10.1503/cmaj.110393] [Medline: 22105750]
18. Vincent C, Davis R. Patients and families as safety experts. CMAJ 2012 Jan 10;184(1):15-16 [FREE Full text] [doi: 10.1503/cmaj.111311] [Medline: 22105747]
19. Emmert M, Sander U, Pisch F. Eight questions about physician-rating websites: a systematic review. J Med Internet Res 2013 Feb 1;15(2):e24 [FREE Full text] [doi: 10.2196/jmir.2360] [Medline: 23372115]
20. Emmert M, Meier F, Pisch F, Sander U. Physician choice making and characteristics associated with using physician-rating websites: cross-sectional study. J Med Internet Res 2013 Aug 28;15(8):e187 [FREE Full text] [doi: 10.2196/jmir.2702] [Medline: 23985220]
21. Maramba ID, Davey A, Elliott MN, Roberts M, Roland M, Brown F, et al. Web-based textual analysis of free-text patient experience comments from a survey in primary care. JMIR Med Inform 2015 May 6;3(2):e20 [FREE Full text] [doi: 10.2196/medinform.3783] [Medline: 25947632]


## Abbreviations

**IGZ:** Health Care Inspectorate (Inspectie voor de Gezondheidszorg)
**OR:** odds ratio
**PRS:** patient rating site


*Edited by G Eysenbach; submitted 19.04.16; peer-reviewed by M Emmert, B Ranard; comments to author 26.05.16; revised version received 09.06.16; accepted 24.06.16; published 15.07.16.*

*Please cite as:*
*Kool RB, Kleefstra SM, Borghans I, Atsma F, van de Belt TH*
*Influence of Intensified Supervision by Health Care Inspectorates on Online Patient Ratings of Hospitals: A Multilevel Study of More Than 43,000 Online Ratings*
*J Med Internet Res 2016;18(7):e198*
*URL: http://www.jmir.org/2016/7/e198/*
*doi:10.2196/jmir.5884*
*PMID:27421302*


©Rudolf Bertijn Kool, Sophia Martine Kleefstra, Ine Borghans, Femke Atsma, Tom H van de Belt. Originally published in the Journal of Medical Internet Research (http://www.jmir.org), 15.07.2016. This is an open-access article distributed under the terms



XSL•FO
RenderX

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



XSL•FO
RenderX

Original Paper

# Investigating the Potential Contribution of Patient Rating Sites to Hospital Supervision: Exploratory Results From an Interview Study in the Netherlands

Sophia Martine Kleefstra[1,2], PhD; Linda C Zandbelt[3], PhD; Ine Borghans[1,4], PhD; Hanneke J.C.J.M de Haes[2], PhD; Rudolf B Kool[4], PhD, MD

[1]Dutch Health Care Inspectorate, Department of Risk Detection and Development, Utrecht, Netherlands

[2]Academic Medical Centre, Department Medical Psychology, University of Amsterdam, Amsterdam, Netherlands

[3]Academic Medical Centre, Department od Clinical and Executive Support, University of Amsterdam, Amsterdam, Netherlands

[4]Radboud Institute for Health Sciences IQ healthcare, Radboud university medical center, Nijmegen, Netherlands

**Corresponding Author:**
Sophia Martine Kleefstra, PhD
Dutch Health Care Inspectorate
Department of Risk Detection and Development
Stadsplateau 1
Utrecht, 3521 AZ
Netherlands
Phone: 31 881205000
Fax: 31 881205001
Email: sm.kleefstra@igz.nl

## Abstract

**Background:** Over the last decades, the patient perspective on health care quality has been unconditionally integrated into quality management. For several years now, patient rating sites have been rapidly gaining attention. These offer a new approach toward hearing the patient's perspective on the quality of health care.

**Objective:** The aim of our study was to explore whether and how patient reviews of hospitals, as reported on rating sites, have the potential to contribute to health care inspector's daily supervision of hospital care.

**Methods:** Given the unexplored nature of the topic, an interview study among hospital inspectors was designed in the Netherlands. We performed 2 rounds of interviews with 10 senior inspectors, addressing their use and their judgment on the relevance of review data from a rating site.

**Results:** All 10 Dutch senior hospital inspectors participated in this research. The inspectors initially showed some reluctance to use the major patient rating site in their daily supervision. This was mainly because of objections such as worries about how representative they are, subjectivity, and doubts about the relevance of patient reviews for supervision. However, confrontation with, and assessment of, negative reviews by the inspectors resulted in 23% of the reviews being deemed relevant for risk identification. Most inspectors were cautiously positive about the contribution of the reviews to their risk identification.

**Conclusions:** Patient rating sites may be of value to the risk-based supervision of hospital care carried out by the Health Care Inspectorate. Health care inspectors do have several objections against the use of patient rating sites for daily supervision. However, when they are presented with texts of negative reviews from a hospital under their supervision, it appears that most inspectors consider it as an additional source of information to detect poor quality of care. Still, it should always be accompanied and verified by other quality and safety indicators. More research on the value and usability of patient rating sites in daily hospital supervision and other health settings is needed.

***(J Med Internet Res 2016;18(7):e201)*** doi:10.2196/jmir.5552

**KEYWORDS**

patient rating sites; patient satisfaction; patient experiences; hospitals; quality of health care; supervision



XSL•FO
RenderX

## Introduction

Over the last decades, the patient's perception of health care quality has been unconditionally integrated into quality management. Traditional patient satisfaction or experience surveys have become accepted tools for measuring health care quality. These tools were demonstrated to add valuable information to professional quality indicators and outcome measures [1,2]. For several years now, a new approach toward hearing the patient's perspective on the quality of health care, by the use of patient rating sites, has rapidly gained attention. These specialized Internet rating sites allow patients to express and rate their experiences and satisfaction with health care providers and institutions. These ratings are intended to be a source of information on quality for other patients looking for health care providers [3-5]. This is especially the case in the United States, Germany, and the United Kingdom where many patients look for information on these sites. Their use as public reporting instrument is often stimulated by governments [6], supporting patients to make explicit comparisons between health care providers, and hereby increasing public accountability and improving quality of care [7-12].

At first, the introduction of patient rating sites caused doctors and policymakers to raise several objections against the use of this information. They were supposed to be vulnerable to a number of pitfalls, such as being manipulated, showing a large variation in the number of ratings for hospitals and physicians, being emotionally burdensome for physicians who were either criticized or even not rated at all, or being biased by selection of patients, for example, by an overrepresentation of dissatisfied patients [5,8,9,11,13,14]. Furthermore, the average number of ratings for individual physicians was still low, implying that the assessments found for physicians may change over time when more patients took part [8]. Subsequently, rating sites are only used by people who have access to and know how to use the Internet, which could cause bias. Finally, information from rating sites was not case-mix adjusted for patient characteristics such as age, level of education, and health status. This is known to be necessary to prevent bias and thus allow the results to be properly interpreted [15,16].

However, recent results from research on rating sites increasingly questioned these arguments and showed certain advantages. Ratings are mostly positive [4,10,11,17,18] and correlate with relevant clinical outcomes such as decreased mortality, readmissions, infection rates, and decubitus [5,8,19-21]. These correlations are at least as strong as for the traditional paper surveys method [9,19,21]. Moreover, in some cases, the real-time nature of rating sites means that feedback can be given rapidly, which might make the information contained on them more up to date and might thus detect episodes of poor care or outliers in a more timely manner than surveys that took place a long time ago [9]. Also, these ratings can be given to all health care professionals and institutions while survey data regard, mostly, only one part of them. Last but not least, there is reason to believe that these rating sites will become commonplace. In fact, an increasing number of people consult the Internet, looking for health care quality information. This rose from 19% of North American adults in 2001 to 88% in 2010; 24% of them consulted review sites. Also, the number of ratings has risen rapidly. In 2010, up to 16% of all US physicians were reviewed [7,8,10,21], whereas 37% of physicians in Germany were reviewed in 2012 [22]. An awareness of 65% of the US population and a usage of 23% shows that patients are increasingly turning to Web-based rating sites [23]. A German study showed that approximately 65% of patients using a rating site have consulted a particular physician based on these ratings [3]. Thus, despite the arguments against the use of rating sites, these sites do have redeeming value that needs to be further explored.

A recent scoping review concluded that although literature about the topic is still limited, social media, and especially patient rating sites, can become a fast and cheap way to gather information about the quality of care and could complement traditional methods [24]. Thus, although some caution interpreting the information is needed, given methodological restrictions [24], using patient rating sites might help to detect poor performance [9,19,21,25,26]. It is therefore stated that neither physicians nor policy makers should underestimate the growing influence of ratings sites for patients in providing information, and for physicians in offering opportunities to improve the quality of their care, based on the concerns mentioned in reviews [3,4].

Due to the potential value of the information for judging the quality of care, some supervisory bodies already use rating sites as an additional source of information [21,27-30]. In England, for example, the Care Quality Commission actively uses patient rating information from the NHS Choices website, alongside other rating sites, to identify potential risks to patient safety [25,28]. Similar initiatives are found in Australia and Ireland [29].

The Dutch health care Inspectorate's (IGZ) supervisory framework for risk detection in hospitals contains in the first place several process and outcome indicators developed to monitor the quality and safety of hospital care [31]. These quality indicators merely focus on clinical care processes and were developed in a collaborative process with the inspectorate, hospital federations, and medical specialist and nursing societies [32]. Furthermore, financial and administrative information, information from calamity reports and earlier visits, and judgments of the inspectorate provide input for risk detection.

Although research shows that IGZ inspectors expect patients to be capable of detecting poor performance or risks that might be missed by regular inspection visits [29], patient's experiences are not yet included, systematically, in Dutch risk detection [33]. However, the inspectorate has become more interested in using information from rating sites to expand their methods to detect poor performance [34] having been stimulated by their colleagues working in health care supervision abroad and by the growing emphasis on patient participation [29]. In addition, an earlier study had already shown that the largest rating site in the Netherlands, ZorgkaartNederland, appeared to be the only social media source that was of additional value for risk-based supervision of elderly care [26]. Using rating sites by the Dutch health care inspectorate to detect poor quality of care could be an important development in several ways. First, the IGZ wants



XSL•FO
RenderX

to involve the patient's perspective in supervision, as the inspectorate's primary client is the citizen [35]. The inspectorate therefore needs reliable sources that express the patient's perspective on quality and safety of health care. Second, stakeholders, such as the health care inspectorate, may give patients a voice by using rating sites, which may encourage them to share their experiences. Besides, it may stimulate health care providers to improve their quality of care, knowing that both patients and stakeholders take these rating sites seriously.

The aim of our study was therefore to explore whether and how patient experiences reported on rating sites can, in the eyes of health care inspectors, contribute to risk identification in hospital care.

We address 3 research questions:

1. Do health care inspectors already use patient experiences on rating sites in their daily supervision of hospitals and in what way?

2. Do inspectors expect patient experiences in hospitals, reported on rating sites, to contribute to their estimation of risk?

3. Does presenting, actively, patient reviews reported on the rating site ZorgkaartNederland alert inspectors in their estimation of risks to patient safety?

## Methods

Given the unexplored nature of the topic, an exploratory, interview study was designed.

We used a semistructured interview approach along with an investigation of the judgment of the review data from a patient rating site. The consolidated criteria for reporting qualitative research (COREQ) guidelines [36] were followed to ensure the completeness of the reporting.

### Sample

For the supervision of hospital care, the IGZ divided the field into 10 segments. Each segment covers 10 hospitals on average with 1 senior inspector being responsible. Our sample thus consisted of 10 senior inspectors.

### Study Design and Procedure

#### *Step 1: First Round of Interviews, Exploring Use and Views*

In January and February 2015, the primary researcher (SK) performed the first round of semistructured interviews with the senior inspectors to establish their actual use in the supervision of health care, of patient experiences reported through rating sites and to explore their views on the potential contribution of such patient ratings (research questions 1 and 2). They were approached by email. The interview guide consisted of general topics concerning attitude to social media in general for working and private purposes; use of patient rating sites for working purposes; and (expected) value of the use of rating sites for supervision. These general topics consisted of several open questions, which were merely explorative: "What do you think of… and why?" Interviews were recorded on audiotape. Field notes were made during the interviews. The interviews lasted up to 1 hour. The first 2 interviews were discussed with 2 researchers (IB and RK) to ensure completeness and interview techniques.

#### *Step 2: Selecting Hospitals and Reviews*

After the first round of interviews, the inspectors were provided with texts of negative reviews on the rating site ZorgkaartNederland regarding one of the hospitals under their supervision. ZorgkaartNederland [37] is the Federation of Patient and Consumer Organizations' (NPCF) noncommercial patient rating site [38]. It has the largest number of patient ratings in the Netherlands, with more than 300,000 ratings in total and 800,000 unique visitors per month. Patients can anonymously rate either the care organization or their care provider on a scale of 1 to 10 based on 6 factors: appointments, accommodation, employees, listening, information, and treatment. The average of the 6 scores yields the overall rating, which is a valid summary of the factor's scores [39]. Patients have to clarify their rating with a written review checked by the website's editorial office. This helps to mediate the risk of unfounded ratings. Even so, the editorial office checks the internet protocol address of every individual review, thus generating information on whether a patient has provided multiple ratings, which could be used to filter out ratings that appear to be duplicates. Patient characteristics are not asked for, so case-mix correction is not possible [40].

We defined a rating as a quantitative score given to a hospital or doctor and a review as a written comment [18]. For each inspector we selected, at random, 1 hospital under their supervision. Only hospitals with at least 50 ratings in the period from November 1, 2013 until October 31, 2014 (1 year) were eligible, to have a substantial number of ratings. Besides, at least 10 negative ratings had to be available for this hospital, as the reviews belonging to these ratings were expected to contain most useful information for inspectorates [26]. Therefore, we categorized the average overall rating using a classification derived from the international known measure of recommendation, the Net Promoter Score. This measure considers the numbers 9 and 10 as positive ("promoters"), the numbers 7 and 8 as neutral, and the numbers 0 till 6 as negative recommendations ("detractors") [41]. If the hospital had less than 10 negative ratings, we selected, at random, another hospital. The hospitals selected had on average 21 negative ratings (see Table 1).



XSL•FO
RenderX

**Table 1.** Rating overview of the hospitals selected and of all the hospitals covered by ZorgkaartNederland (November 1, 2013-31, October 2014).

| | Ratings of 10 selected hospitals (mean of the 10 hospitals (range)) | Ratings of all (94) hospitals on ZorgkaartNederland (mean of the 94 hospitals (range)) |
|---|---|---|
| Total number of ratings | 129 (65-170) | 173 (4-859) |
| Mean rating score[a] | 8.2 (7.9-8.6) | 8.5 (7.5-9.1) |
| Positive ratings (score>8.4) | 86 (40-116) | 122 (3-598) |
| Neutral ratings (score 6.5-8.4) | 22 (11-33) | 36 (1-250) |
| Negative ratings (score<6.5) | 21 (12-28) | 15 (0-56) |
| Percentage >6.4 | 83.6 (78.5-89.7) | 91.1 (67.9-100) |
| Percentage <6.5 | 16.4 (10.3-21.5) | 8.9 (0-32.1) |

[a] Rating score: average of 6 scores on a scale of 1 to 10 regarding appointments, accommodation, employees, listening, information, and treatment.

Subsequently, we presented the texts of the negative reviews of the hospital selected in an Excel sheet, which was sent by email to the inspectors. We also provided the hospital's contextual information such as the name, the mean rating, the total number of positive and negative ratings, and the percentage of negative ratings, as compared with other hospitals, and what level the review was attributed to: hospital, location, department, or doctor. Inspectors were asked to score the relevance of each negative review for the health care inspectorate according to a previously developed ordinal assessment scheme [26]: "no additional value (0)," "relevant, information leads to a signal in the file of the organization (1)," "relevant, information leads to further investigations (2)," or "relevant, information leads to immediate action (3)." We choose to reveal the name of the selected hospital to explore whether inspectors would find out new information or merely information that was supportive of what they already knew from their experiences with the hospital. Inspectors filled in the score list and returned it to the researcher before the second interview.

### *Step 3: Scoring Negative Reviews and Identifying Underlying Motives*

The primary researcher (SK) performed a second round of interviews from April until June 2015. The aim was to determine whether the reviews contained information on risks to patient safety (research question 3). These interviews consisted of 2 parts. In part 1, inspectors were queried about their judgment of each negative review and were asked what elements in the text of the reviews triggered their scoring. We provided some possible triggers, such as the subject, the tone, the concreteness, or the extensiveness of the review. In addition to these, the inspectors could always add new triggers. In part 2, inspectors were asked their general opinion about the use and value of the judged reviews for daily supervision work. The topic list included items such as usability, reliability, new or known information, and value for risk estimation. These interviews were also tape-recorded.

### Analysis

All interviews were transcribed verbatim and were sent to the interviewees for triangulation.

They were analyzed following guidelines for qualitative research [42,43] and by using a digital qualitative data analysis program, Atlas-ti [44]. Analysis was performed in parallel with the interviewing. In the first round of interviews, the first interviews were analyzed inductively, aiming to explore and identify relevant views and propositions. In the second round of interviews, open coding (summarizing and categorizing the data) was gradually replaced with axial coding (confirmation of codes and the identification of broader relationships). Finally, data were clustered across interviews to derive common themes. The inspector's scoring of the negative reviews was analyzed descriptively [26]. The arguments for the scores were described. We performed a member check by sending all quotes to all interviewees to ensure interpretation and hereby validity.

## Results

### Sample

All 10 senior inspectors consented to participate in both interviews. Their average age was 53 years (range 40-64). Seven were women. All inspectors were educated as a health care professional and had worked in a hospital for several years. The average number of working years as an inspector was 8.5 years (range 1-17). Four inspectors used social media (Twitter, Facebook) for private purposes. All used the Internet for their work (Google, ZorgkaartNederland, Twitter, news websites).

### Inspector's Current Use of Patient Rating Sites in Daily Supervision

The first round of interviews addressed the first research question, whether health care inspectors already used patient experiences on rating sites in their daily supervision of hospitals and in what way.

Seven inspectors used ZorgkaartNederland to gather information in their supervision work. When preparing their annual meeting with the board of a hospital or in case of reports of serious incidents, they looked for information on search machines such as Google and then ended up at the patient rating site ZorgkaartNederland.

> *Then I google that person. You end up at ZorgkaartNederland very quickly. The first hit of Google apparently is ZorgkaartNederland.* [Respondent 3]

> *In particular I use ZorgkaartNederland, in any case I look at it in preparation for the annual board interview. And, if we focus on a specific doctor*



XSL•FO
RenderX

> *involved in a report or for example because of the suspicion of incompetence, then I check ZorgkaartNederland for the individual judgment relating to the doctor.* [Respondent 2]

Three inspectors did not use the patient rating site, ZorgkaartNederland. They did, however, gather their information from the Internet, but in their cases from hospital websites, newsletters, or news websites, not from a source that contains the patient's perspective.

> *I read newsletters from hospitals. (…) But Twitter is also a possible source. (…) For me that is easy to read, and very handy because I can scan very quickly whether it is valuable for me or not.* [Respondent 6]

## Inspector's Anticipated Value of Patient Rating Sites for Daily Supervision

The first round of interviews also addressed the second research question, whether health care inspectors expect patient experiences in hospitals, reported on rating sites, to contribute to their estimation of risk to patient safety.

All inspectors who ended up at ZorgkaartNederland indicated that they find it hard to use this information or give weight to this information in their daily supervision.

> *I think you should be very careful with this information. It must be seen as a signal, not more than that. A signal deserves to be taken seriously and to be properly checked and verified.* [Respondent 10]

> *What do you do with it? You take it with you. In that way you use it, but concretely in the conversation with the hospital board, or, in the reports, no, you do not use it that way.* [Respondent 1]

Thus, apart from a source for gathering information, the 7 inspectors using ZorgkaartNederland did not apply the content of the information for risk identification in their daily supervision practice. However, they saw the reviews as a signal, providing interesting background or contextual information. In the opinion of 5 inspectors, these signals should always be verified and checked by other available information.

> *In fact it is an indicator. An indicator always needs further research. It must be seen in combination with other indicators: what are the connections and the relevant themes?* [Respondent 2]

The inspectors brought up 3 main doubts concerning the weight and value of ZorgkaartNederland as a source for identifying risks. Firstly, 4 inspectors feared bias or selectivity, that is they felt that only a small group of people uses rating sites.

> *The number of reviews is too small to be taken seriously. Only a small group of patients makes the effort.* [Respondent 5]

Inspectors felt that this group is probably not representative of the patient population of a hospital. For example, hospitals might stimulate very satisfied patients to rate their experiences, to raise their average rating. Besides, positive reviews may have been posted by family and friends of the doctor. Second, 9 inspectors indicated that reviews are often too subjective and emotionally driven. Accordingly, reviews may polarize public opinion at a certain moment and can be used for unnecessarily blaming the doctor.

> *I feel the psychology of reviewers on a rating site is interesting. In fact, there is a lot of psychology on those sites. People parrot each other easily and therefore strengthen the message and are thus polarizing what happened at a certain moment. And that gives an incorrect picture of the hospital or doctor. It is influenced too much by the moment and the polarization. We should be aware of that.* [Respondent 8]

> *It can be used for blaming and shaming. That is very easy on the Internet because it is safe and anonymous.* [Respondent 9]

Third, inspectors had doubts about the relevance of the content of reviews for the inspectorate's estimation of risk. Negative reviews were thought to contain mostly remarks on the way patients are addressed, the bad food, signage, or waiting times, not about potential risks to safety.

> *Patients talk on a very basic level, often about how patients are addressed, and that is not within our remit.* [Respondent 5]

> *I do not know how to interpret the reviews. You know, if a doctor is nice he gets an eight although technically speaking he is not so good. The patient cannot interpret that. (…) I feel that is no use for me.* [Respondent 6]

If information on ZorgkaartNederland could be integrated into other sources of information on patient safety, most inspectors would consider this information to contribute toward the identification of risks. They indicated that the value of reviews for their supervision would improve if the reviews were supported by facts and were substantial but also that the tone of the texts matters.

> *It depends on whether the review is supported by facts. If it is written in concrete, correct sentences (…) I would rather adopt it than when it is a story of verbal abuse like "it was really awful" with a lot of emotions.* [Respondent 10]

Yet, the inspectors indicated that they would be triggered to act if a review contains medical errors, serious incidents, damage, unacceptable care or, shortcomings of care. Those reviews would be taken more seriously than reviews about how patients are addressed or about complaints. They would also pay attention when the number of negative reviews suddenly rises because this could be a signal of failing. The inspector should have the feeling that the review was not written impulsively,

> *"but that another reasonable patient could echo this judgment as well.* [Respondent 8]

## Experienced Relevance of Patient Rating Sites for Daily Supervision

The second round of interviews addressed the third research question whether actively presenting patient reviews reported



XSL•FO
RenderX

on the rating site ZorgkaartNederland alerts inspectors in their estimation of risks to patient safety.

In total, 207 negative reviews were presented to the inspectors, who scored these according to their relevance. The inspectors scored 47 (22.7%) reviews "relevant" (score 1, 2, or 3; see Table 2).

**Table 2.** The relevance of reviews as scored by the inspectors.

| | Negative reviews (N) | Percentage | Percentage of "relevant" scores |
|---|---|---|---|
| No additional value (0) | 160 (in 10 hospitals) | 77.3 | — |
| Relevant, information leads to a signal in the file of the organization (1) | 31 (in 7 hospitals) | 15.0 | 66.0 |
| Relevant, information leads to further investigations (2) | 15 (in 6 hospitals) | 7.2 | 31.9 |
| Relevant, information leads to immediate action (3) | 1 (in 1 hospital) | 0.5 | 2.1 |
| Total | 207 | 100 | 100 |

### *The Reasons Reviews Were Considered Irrelevant*

Most of the reviews that were scored as nonrelevant for supervision (160/207) were labeled as a complaint dealing with how patients were addressed, the attitude of the doctor, information and communication, or waiting times. Inspectors indicated that dealing with such complaints is a task of the hospital itself, that is, the board or a complaint officer or committee.

> *This is about how the patient is addressed such as bad experiences with being listened to. I reckon that this happens in every hospital and I am convinced that a lot of improvements can be made in this respect, but it is not a task of the health care inspectorate.* [Respondent 7]

Other motives not to score the review as relevant were their vagueness, the shortness of the description, or the highly emotional tone such as with comments like:

> *"He is a horrible man." That man may well be horrible, but what can the health care inspectorate do about it?* [Respondent 3]

### *The Reasons Reviews Were Considered Relevant*

Thirty-one reviews (31/207; 15%) were scored as "relevant, information leads to a signal in the file of the organization" (score 1). The reasons why inspectors gave this score were:

- The review mentioned risks concerning quality and safety.
- The review had a medical content.
- The review could indicate a structural problem, such as shortcomings in care for vulnerable elderly patients or children; therefore, it could contribute to the compilation of a file on that particular hospital or department.
- The doctor was also an instructor to students.
- The department or doctor were well-known, for instance from an earlier investigation, or an underperforming department.

> *I know this doctor, he came up more often in conversations. He is also mentioned in an earlier investigation. Although no serious incidents have been reported against him, he is known to be a difficult man to deal with—so to speak!* [Respondent 1]

Fifteen reviews (15/207; 7.2%) were scored as "relevant, information leads to further investigations" (score 2). The reasons the inspectors gave for considering these reviews to be of greater relevance were medical, procedural, or related to the hospital's profile:

- The review mentioned serious incidents or surgical or medical errors, complications, or damage to the patient or other major consequences such as a long length of stay; or the review concerned medication, it was, for instance, forgotten, or a prescription meant for another patient was given in error during discharge from the hospital.
- The review concerned actual procedural themes in the hospital, for instance, deficiencies in procedures concerning the primary treating physician, about cardiac rehabilitation, or about shortcomings with anticoagulants.

If reviews concerned the hospital's profile, this might indicate 2 possibilities. Either the review was about a topic in which the hospital was not specialized;

> *This hospital has no department for genetic research, so in that context, if genetic factors play a role, it should be taken care of by specific procedures. And, according to this review there was insufficient attention given to genetic factors.* [Respondent 8]

Or the review was related to a topic in which the hospital was specialized.

> *This hospital is a bariatric center. Given that context this should not have happened here.* [Respondent 2]

One review (1/207; 0.5%) was indicated as "relevant, information leads to immediate action" (score 3). The considerations given by the inspector were:

- The review described a serious incident, which was also reported to the inspectorate.
- The review concerned an already notorious doctor.
- Moreover, the hospital had not reacted properly after this serious incident.



XSL•FO
RenderX

### *Additional Considerations Regarding Relevance of Reviews*

The inspectors mentioned several other considerations for judging reviews to be of greater relevance:

- the number of reviews concerning a specific department, doctor, or topic;
- the concreteness of the review;

> *Five operations, two times outpatient operations, five infections; these are concrete facts which make me wonder what kind of operation room was that?* [Respondent 4]

- their own opinion and experiences with how the hospital was functioning;

> *I am aware of a serious incident that happened recently in this department, so when I saw this review I was alerted. Then I saw another review about a doctor and again it was this same department. So maybe there is more going on there.* [Respondent 3]

- the given period of time and the actual events that took place in the hospital;

> *This hospital has had a lot of negative publicity in that specific period. I think that is reflected in the negative reviews.* [Respondent 7]

- the ranking of the hospital on other well-known ranking lists;

> *Since several years this hospital is on top of a number or ranking lists. However, last year it fell down (…) I think it is interesting to interpret this period, especially where does this organization come from, where are they now and where are they heading for?* [Respondent 2]

- what was already known by the IGZ from other quality indicators;
- the contextual information about the mean scores of all hospitals was considered by most inspectors in their assessment of the reviews as valuable, but never decisive.

> *The percentage of negative reviews is high compared to other hospitals, but maybe this hospital challenges patients to offer a rating on ZorgkaartNederland. That fits in with the positive picture I have of this hospital.* [Respondent 3]

### New Information or Already Known?

The actions of 9 inspectors were triggered especially by reviews that confirmed their knowledge about, and experience with, the hospital. In these instances, the reviews on ZorgkaartNederland supported the other sources of information used. Five inspectors explicitly indicated that the reviews rendered new information, mostly concerning a specific doctor or department that was mentioned more than once in the reviews.

> *For me it resulted in two new points of attention: this doctor, who was mentioned four times and I have never heard of, and also the critical remarks about that specific department I did not know of.* [Respondent 5]

In summary, after having been confronted with the reviews, the inspectors mentioned 2 ways in which they could use this information from ZorgkaartNederland in future supervisory work. According to 9 inspectors, this information could be used to put topics, departments, or specific doctors onto the agenda in the yearly interview with a hospital board.

> *I would mention it as a signal: I saw on ZorgkaartNederland that...Have you seen it as well and what do you think about it? And if so, what have you done about it?* [Respondent 10]

Three inspectors indicated that this new information could be used in unannounced visits to the hospital, especially referring to specific departments who came to attention through the reviews.

> *We assess a lot of things, indicators, reports of serious incidents, but if you look for themes in order to make an unannounced visit, this could be part of it, definitely. People make an effort to write a review on ZorgkaartNederland, they do that on purpose.* [Respondent 9]

## Discussion

We examined whether and how patient experiences as reported on patient rating sites have a potential to contribute to hospital inspectors identification of risks to safety. Currently, most inspectors only use patient experiences on the patient rating site, ZorgkaartNederland, as a source for gathering background or contextual information about a hospital or a doctor. It automatically arises with searching the Internet. However, for most inspectors, this appears to lead to the question: what exactly to do with the ratings and reviews and how to determine the value of the picture they get? This could be caused by 3 main objections brought up by the inspectors at the beginning of this study. First, inspectors worry about how representative the patient rating sites are, given, for instance, the selected group of patients responding and the relatively low number of ratings. Second, they indicate that reviews are often too subjective and emotionally driven. Third, they had doubts about the relevance of the content of these reviews for supervision.

Earlier research showed, too, another objection among inspectors to the use of patient rating sites for supervision. This was their concern about whether patients are able to evaluate the medical expertise and capabilities of an individual doctor [29,45].

Concerning how far rating sites are representative, it is known from literature that users of patient rating sites significantly differ from nonusers on sociodemographic and psychographic variables and health status. Users are significantly younger and more highly educated. Also, female patients and patients with chronic diseases use patient rating sites more often than other patient groups [17,46]. However, research on the data provided by ZorgkaartNederland [40] showed that the self-selected sample of patients on ZorgkaartNederland did, in fact, lead to representative ratings about Dutch health care in hospitals. Moreover, research into the content of reviews showed that the



XSL•FO
RenderX

review process is not just a one-off reflection of a single moment but contextualizes this within a series of previous experiences [45]. This may place the prevalent "n=1" objection in perspective.

The subjectivity of patient's assessment is a well-known discussion in literature. Indeed, a patient's assessment of care is subjective, by nature. Nevertheless, a lot of research has been done, showing positive relationships between patient's (subjective) assessments and the quality of care, patient safety, and clinical effectiveness [47-53].

Although there is evidence of the correlation between scores on patient rating sites and quality indicators and clinical outcomes on a hospital level [5,8,19-21], little research has been carried out on the association between patient ratings and physician quality metrics. Gao et al found a significant positive relationship between Web-based ratings and physician quality as shown by board certification, education, and malpractice claims [8]. However, more research on this topic is needed to overcome this objection.

Despite their reservations regarding the use of patient rating sites for daily supervision, when confronted with the text of negative reviews from one of the hospitals under their supervision, inspectors scored 23% of the reviews as being relevant for risk estimation. Reviews were indicated as relevant when they contained information about major safety problems such as medication errors, serious incidents, severe damage or consequences for the patient, structural organizational problems such as a malfunctioning department or doctor, actual themes, and whether the reviews are in line with the hospital's profile. Many of these "medical" indicators of possible relevance were also mentioned by inspectors at the beginning of the study, before having scored the reviews presented. However, the scoring of the reviews also revealed new relevant indicators such as structural and procedural organizational problems, which could produce a relevant score for risk estimation.

Compared to previous research carried out on reviews from ZorgkaartNederland concerning their additional value for supervision in the long-term elderly care [26], the percentage of reviews considered relevant by hospital inspectors was lower (23% vs 62%). However, from the relevant hospital reviews, 34% is seen as "relevant, information leads to further investigations" (score 2) or even "relevant, information leads to immediate action" (score 3), compared to 15% in the elderly care. As compared with long-term elderly care, safety issues in hospitals might be judged as being serious at an earlier stage, given the high-risk processes involved. The high number of reviews judged to identify safety issues is in line with patient safety literature, which states that there is evidence to suggest that hospital patients can be used as partners in identifying poor and unsafe practice and help enhance effectiveness and safety [48,54]. Although most comments are classified as physician-related concerns [4,14], content analyses of reviews in literature showed 3 dominant themes indicated by patients: interpersonal manner, technical competence, and system issues. These all include potential risks to patient safety [14]. It is important to note that the use of rating sites is likely to increase in the near future when the generation socialized with social media reaches the age in which health questions and doctors become dominant. As a result, these kinds of sources might become even more relevant [3,4,12,46] for patients and physicians, as well as for stakeholders such as the health care inspectorate.

### Implications and Future Research

Hospital inspectors at first showed some restraint in their concrete use of ZorgkaartNederland in their daily supervision. However, after being confronted, the negative reviews of one of the hospitals under their supervision, most inspectors were cautiously positive about the contribution of the reviews to their risk identification. Nevertheless, they insisted that the use of rating sites should always be accompanied and verified by clinical indicators. The caution of inspectors for the use of reviews from patients is a point of concern for supervision policy in the near future. It appears to be worthwhile to provide health care inspectors regularly with a summary of negative reviews on carefully edited rating sites such as ZorgkaartNederland, complemented with contextual information, regarding hospitals under their supervision. Almost all inspectors indicated that specific themes, departments, or doctors on ZorgkaartNederland could be presented in their annual interview with the hospital board. Also, specific departments that showed up negatively in the reviews could be subjected to unannounced visits. However, evaluating the value and usability of this additional source for hospital supervision in the near future is necessary. Furthermore, it takes more research to understand and support the additional value of the patient's perspective on quality of health care, for instance, by comparing the patient's perspective with clinical outcome indicators or with supervision judgements.

A positive aspect of using ratings and reviews in supervision is the availability of actual information, in addition to the yearly available conventional quality indicators. Thus, a more efficient way of risk-based prioritizing within a huge number of health care organizations is a possibility [26]. This is especially important in health care sectors with a substantial number of organizations or professionals such as the elderly care sector, general practitioners, dentists, and pharmacists. In this way, patient ratings and reviews can become a structural part of the supervisory framework for risk detection.

However, most of the ratings on ZorgkaartNederland are positive, as is the case for most rating sites [4,10,11,17,18]. Furthermore, the percentage of negative ratings is decreasing in time, from 19.9% in 2010 to 7.2% in 2015 [55]. This may implicate that poor performance cannot be exclusively depicted by rating sites. Preferably, information from rating sites should be accompanied by other sources to express the patient's perspective, such as general patient experiences or satisfaction surveys. Furthermore, patients could be stimulated by the government, hospitals, health care providers, and patient organizations to place their experiences on rating sites such as ZorgkaartNederland, to cover a more broad spectrum of patient experiences. For example, the branch organization of long-term elderly care and the NPCF, as owner of ZorgkaartNederland, organizes so-called road teams since 2015. These teams visit institutions of elderly care with mobile devices connected to the elderly care section of ZorgkaartNederland, interviewing



XSL•FO
RenderX

clients and relatives to increase the number of ratings substantially. Moreover, since July 1, 2014 a National Reporting Centre for Health Care Complaints (Landelijk Meldpunt Zorg) in the Netherlands gives patients and relatives an opportunity to express their complaints about care, always after having first complained at their provider. The health care inspectorate is given insight in these complaints and can use this information as additional source to detect poor performance from the patient's perspective.

### Strength and Limitations

This study has strengths and also limitations. The fact that the patient rating site ZorgkaartNederland is an independent, noncommercial website, with its own editorial office that judges the reviews one by one on their substantiating text and checks on the sender of the rating, is a strength of this patient rating site. It increases the value of the reviews. This is not necessarily the case with all patient rating sites in other countries.

The hospitals selected were not necessarily representative of hospitals on ZorgkaartNederland. However, the focus of our research was on the identification of risks in the texts of the negative reviews. Therefore, we wanted a substantial number of negative reviews per hospital and put the minimum threshold on 10. In that way, it was possible to identify trends, themes, departments, or doctors that were, for instance, mentioned more than once.

In this research design, we selected, for each inspector, a hospital for which he or she was responsible. In fact, most inspectors have known these hospitals for some time. They therefore assess the reviews according to their own point of reference, consisting of their accumulated knowledge and experiences. This can be a support to information already known by the inspectors, for instance, about a dysfunctional department. However, this could also blind the inspector to new insights or safety aspects. It would be worthwhile to investigate, in a future study, whether an inspector unacquainted with a certain hospital, would come to the same or a different selection of relevant reviews.

Furthermore, this is a case study among hospital inspectors in the Dutch health care setting, and more research in other settings is needed to draw general conclusions about the usability of patient rating sites for risk detection in supervision.

### Conclusions

Patient rating sites may contribute to the risk-based supervision of hospital care of a health care inspectorate. Health care inspectors do have several objections against the use of patient rating sites for daily supervision. However, when they are presented with texts of negative reviews from a hospital under their supervision, it appears that most inspectors consider it as an additional source of information from the patient's perspective to detect poor quality of care. Still, it should always be accompanied and verified by other quality and safety indicators. Preferably, it should also be accompanied by other methods to reveal patient's experiences, to broaden the patient's perspective on quality and safety of care. Furthermore, more research on the value and usability of patient rating sites in daily hospital supervision and other health care settings is needed.

### Acknowledgments

The authors thank the NPCF for giving full access to the data of ZorgkaartNederland.

### Conflicts of Interest

None declared.

### References

1. Arah OA, ten Asbroek AH, Delnoij DM, de Koning JS, Stam PJ, Poll AH, et al. Psychometric properties of the Dutch version of the Hospital-level Consumer Assessment of Health Plans Survey instrument. Health Serv Res 2006 Feb;41(1):284-301 [FREE Full text] [doi: 10.1111/j.1475-6773.2005.00462.x] [Medline: 16430612]
2. Kleefstra SM, Kool RB, Veldkamp CM, Winters-van der Meer AC, Mens MA, Blijham GH, et al. A core questionnaire for the assessment of patient satisfaction in academic hospitals in The Netherlands: development and first results in a nationwide study. Qual Saf Health Care 2010 Oct;19(5):e24. [doi: 10.1136/qshc.2008.030825] [Medline: 20577001]
3. Emmert M, Meier F, Pisch F, Sander U. Physician choice making and characteristics associated with using physician-rating websites: cross-sectional study. J Med Internet Res 2013;15(8):e187 [FREE Full text] [doi: 10.2196/jmir.2702] [Medline: 23985220]
4. Emmert M, Meier F, Heider AK, Dürr C, Sander U. What do patients say about their physicians? an analysis of 3000 narrative comments posted on a German physician rating website. Health Policy 2014 Oct;118(1):66-73. [doi: 10.1016/j.healthpol.2014.04.015] [Medline: 24836021]
5. Strech D. Ethical principles for physician rating sites. J Med Internet Res 2011;13(4):e113 [FREE Full text] [doi: 10.2196/jmir.1899] [Medline: 22146737]
6. Care Quality Commission. People's experience of care: What we want to know and why URL: http://www.cqc.org.uk/node/2112 [accessed 2016-06-17] [WebCite Cache ID 6iKWftnCQ]
7. Emmert M, Sander U, Pisch F. Eight questions about physician-rating websites: a systematic review. J Med Internet Res 2013;15(2):e24 [FREE Full text] [doi: 10.2196/jmir.2360] [Medline: 23372115]



XSL•FO
RenderX

8. Gao GG, McCullough JS, Agarwal R, Jha AK. A changing landscape of physician quality reporting: analysis of patients' online ratings of their physicians over a 5-year period. J Med Internet Res 2012;14(1):e38 [FREE Full text] [doi: 10.2196/jmir.2003] [Medline: 22366336]
9. Greaves F, Pape UJ, King D, Darzi A, Majeed A, Wachter RM, et al. Associations between Internet-based patient ratings and conventional surveys of patient experience in the English NHS: an observational study. BMJ Qual Saf 2012 Jul;21(7):600-605. [doi: 10.1136/bmjqs-2012-000906] [Medline: 22523318]
10. Kadry B, Chu LF, Kadry B, Gammas D, Macario A. Analysis of 4999 online physician ratings indicates that most patients give physicians a favorable rating. J Med Internet Res 2011;13(4):e95 [FREE Full text] [doi: 10.2196/jmir.1960]
11. 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]
12. Reimann S, Strech D. The representation of patient experience and satisfaction in physician rating sites. A criteria-based analysis of English- and German-language sites. BMC Health Serv Res 2010;10:332 [FREE Full text] [doi: 10.1186/1472-6963-10-332] [Medline: 21138579]
13. Galizzi MM, Miraldo M, Stavropoulou C, Desai M, Jayatunga W, Joshi M, et al. Who is more likely to use doctor-rating websites, and why? A cross-sectional study in London. BMJ Open 2012;2(6):e001493 [FREE Full text] [doi: 10.1136/bmjopen-2012-001493] [Medline: 23148340]
14. López A, Detz A, Ratanawongsa N, Sarkar U. What patients say about their doctors online: a qualitative content analysis. J Gen Intern Med 2012 Jun;27(6):685-692 [FREE Full text] [doi: 10.1007/s11606-011-1958-4] [Medline: 22215270]
15. Emmert M, Sander U, Esslinger AS, Maryschok M, Schöffski O. Public reporting in Germany: the content of physician rating websites. Methods Inf Med 2012;51(2):112-120. [doi: 10.3414/ME11-01-0045] [Medline: 22101427]
16. Hekkert KD, Cihangir S, Kleefstra SM, van den Berg B, Kool RB. Patient satisfaction revisited: a multilevel approach. Soc Sci Med 2009 Jul;69(1):68-75. [doi: 10.1016/j.socscimed.2009.04.016] [Medline: 19446942]
17. Burkle CM, Keegan MT. Popularity of internet physician rating sites and their apparent influence on patients' choices of physicians. BMC Health Serv Res 2015;15:416 [FREE Full text] [doi: 10.1186/s12913-015-1099-2] [Medline: 26410383]
18. Ellimoottil C, Hart A, Greco K, Quek ML, Farooq A. Online reviews of 500 urologists. J Urol 2013 Jun;189(6):2269-2273. [doi: 10.1016/j.juro.2012.12.013] [Medline: 23228385]
19. Bardach NS, Asteria-Peñaloza R, Boscardin WJ, Dudley RA. The relationship between commercial website ratings and traditional hospital performance measures in the USA. BMJ Qual Saf 2013 Mar;22(3):194-202 [FREE Full text] [doi: 10.1136/bmjqs-2012-001360] [Medline: 23178860]
20. Greaves F, Pape UJ, King D, Darzi A, Majeed A, Wachter RM, et al. Associations between Web-based patient ratings and objective measures of hospital quality. Arch Intern Med 2012 Mar 12;172(5):435-436. [doi: 10.1001/archinternmed.2011.1675] [Medline: 22331980]
21. Rozenblum R, Bates DW. Patient-centred healthcare, social media and the internet: the perfect storm? BMJ Qual Saf 2013 Mar;22(3):183-186. [doi: 10.1136/bmjqs-2012-001744] [Medline: 23378660]
22. Emmert M, Halling F, Meier F. Evaluations of dentists on a German physician rating Website: an analysis of the ratings. J Med Internet Res 2015;17(1):e15 [FREE Full text] [doi: 10.2196/jmir.3830] [Medline: 25582914]
23. Hanauer DA, Zheng K, Singer DC, Gebremariam A, Davis MM. Public awareness, perception, and use of online physician rating sites. JAMA 2014 Feb 19;311(7):734-735. [doi: 10.1001/jama.2013.283194] [Medline: 24549555]
24. Verhoef LM, Van de Belt TH, Engelen LJ, Schoonhoven L, Kool RB. Social media and rating sites as tools to understanding quality of care: a scoping review. J Med Internet Res 2014;16(2):e56 [FREE Full text] [doi: 10.2196/jmir.3024] [Medline: 24566844]
25. Greaves F, Ramirez-Cano D, Millett C, Darzi A, Donaldson L. Harnessing the cloud of patient experience: using social media to detect poor quality healthcare. BMJ Qual Saf 2013 Mar;22(3):251-255. [doi: 10.1136/bmjqs-2012-001527] [Medline: 23349387]
26. Van de Belt TH, Engelen LJ, Verhoef LM, van der Weide MJ, Schoonhoven L, Kool RB. Using patient experiences on Dutch social media to supervise health care services: exploratory study. J Med Internet Res 2015;17(1):e7 [FREE Full text] [doi: 10.2196/jmir.3906] [Medline: 25592481]
27. Framework for Public & Service User Involvement in Health and Social Care Regulation in Ireland. Dublin: Health and Social Care Regulatory Forum; Dec 2009:1-40.
28. Care Quality Commission. Voices into action. How the Care Quality Commission is going to involve people URL: http://web.archive.org/web/20101214155613/http://cqc.org.uk/_db/_documents/A4_Report_2009_01.pdf [accessed 2016-07-13] [WebCite Cache ID 6iy7ACNjb]
29. Adams SA, van de Bovenkamp H, Robben P. Including citizens in institutional reviews: expectations and experiences from the Dutch Healthcare Inspectorate. Health Expect 2015 Oct;18(5):1463-1473. [doi: 10.1111/hex.12126] [Medline: 24112128]
30. Grabosky P. Beyond Responsive Regulation: The expanding role of non-state actors in the regulatory process. Regulation & Governance 2013 Mar;7(1):114-123. [doi: 10.1111/j.1748-5991.2012.01147.x]
31. Inspectie voor de Gezondheidszorg. 2014. Basisset Kwaliteitsindicatoren Ziekenhuizen 2014 URL: http://www.igz.nl/zoeken/document.aspx?doc=Basisset+Kwaliteitsindicatoren+Ziekenhuizen+2014&docid=6076 [accessed 2016-07-13]




XSL•FO
RenderX

32. Oldenburg CS, van den Berg JM, Leistikow IP, Dutch Health Care Inspectorate. BMJ Outcomes (Inaugural collection). 2015. All aboard, getting nationwide indicators on the rails URL: http://outcomes.bmj.com/index.php/journal/all-aboard-getting-nationwide-indicators-on-the-rails [accessed 2016-06-24] [WebCite Cache ID 6iV3ph92C]
33. Robben P. Patient's perpective of the Dutch healthcare Inspectorate. In: Burgerperspectief van de IGZ. Utrecht: Dutch Healthcare Inspectorate; 2011.
34. Drewes H, Huisman A, van de Belt T, Schoonhoven L, Engelen L, Kool RB. Toezicht in de 21e eeuw. Zicht op kwaliteit van zorg door sociale media? KIZ, tijdschrift over kwaliteit en veiligheid in de zorg 2013;2:28-30.
35. Inspectie voor de Gezondheidszorg. 2016 Jan. Meerjaren-beleidsplan Gezond vertrouwen 2016-2019 URL: http://www.igz.nl/zoeken/download.aspx?download=IGZ+Meerjarenbeleidsplan+2016-2019.pdf [accessed 2016-07-13] [WebCite Cache ID 6iyBwDnS0]
36. Tong A, Sainsbury P, Craig J. Consolidated criteria for reporting qualitative research (COREQ): a 32-item checklist for interviews and focus groups. Int J Qual Health Care 2007 Dec;19(6):349-357 [FREE Full text] [doi: 10.1093/intqhc/mzm042] [Medline: 17872937]
37. ZorgkaartNederland. URL: https://www.zorgkaartnederland.nl/ [accessed 2015-12-04] [WebCite Cache ID 6dWPrZtve]
38. Patiëntenfederatie Nederland. URL: https://www.npcf.nl/ [accessed 2012-04-25] [WebCite Cache ID 6iV425xle]
39. Krol MW, de Boer D, Rademakers JJ, Delnoij DM. Overall scores as an alternative to global ratings in patient experience surveys; a comparison of four methods. BMC Health Serv Res 2013;13:479 [FREE Full text] [doi: 10.1186/1472-6963-13-479] [Medline: 24245726]
40. Geesink R. Increasing importance of patient ratings. A study on the validity and reliability of ZorgkaartNederland.nl. Amsterdam: Vrije Universiteit Amsterdam, Policy and Organization of Healthcare; Jun 2013:1-47.
41. Reichheld FF. The one number you need to grow. Harv Bus Rev 2003 Dec;81(12):46-54, 124. [Medline: 14712543]
42. Boeije H. Analyseren in kwalitatief onderzoek. Den Haag: Boom Lemma Uitgevers; 2005.
43. Mayring P. Qualitative Content Analysis. Forum: Qualitative Social Research/Forum: Qualitative Sozialforschung 2000;1(2):Art20 [FREE Full text]
44. Atlas-ti Scientific Software Development GmbH. Berlin; 2015. URL: http://atlasti.com/ [accessed 2016-07-14] [WebCite Cache ID 6izPIr6Md]
45. Adams SA. Sourcing the crowd for health services improvement: The reflexive patient and “share-your-experience” websites. Soc Sci Med 2011 Apr;72(7):1069-1076. [doi: 10.1016/j.socscimed.2011.02.001] [Medline: 21414701]
46. Terlutter R, Bidmon S, Röttl J. Who uses physician-rating websites? Differences in sociodemographic variables, psychographic variables, and health status of users and nonusers of physician-rating websites. J Med Internet Res 2014;16(3):e97 [FREE Full text] [doi: 10.2196/jmir.3145] [Medline: 24686918]
47. Boulding W, Glickman SW, Manary MP, Schulman KA, Staelin R. Relationship between patient satisfaction with inpatient care and hospital readmission within 30 days. Am J Manag Care 2011 Jan;17(1):41-48 [FREE Full text] [Medline: 21348567]
48. Doyle C, Lennox L, Bell D. A systematic review of evidence on the links between patient experience and clinical safety and effectiveness. BMJ Open 2013;3(1):1-19 [FREE Full text] [doi: 10.1136/bmjopen-2012-001570] [Medline: 23293244]
49. Glickman SW, Boulding W, Manary M, Staelin R, Roe MT, Wolosin RJ, et al. Patient satisfaction and its relationship with clinical quality and inpatient mortality in acute myocardial infarction. Circ Cardiovasc Qual Outcomes 2010 Mar;3(2):188-195 [FREE Full text] [doi: 10.1161/CIRCOUTCOMES.109.900597] [Medline: 20179265]
50. Isaac T, Zaslavsky AM, Cleary PD, Landon BE. The relationship between patients' perception of care and measures of hospital quality and safety. Health Serv Res 2010 Aug;45(4):1024-1040 [FREE Full text] [doi: 10.1111/j.1475-6773.2010.01122.x] [Medline: 20528990]
51. Jaipaul CK, Rosenthal GE. Do hospitals with lower mortality have higher patient satisfaction? A regional analysis of patients with medical diagnoses. Am J Med Qual 2003;18(2):59-65. [Medline: 12710554]
52. Luxford K. What does the patient know about quality? Int J Qual Health Care 2012 Oct;24(5):439-440 [FREE Full text] [doi: 10.1093/intqhc/mzs053] [Medline: 22918180]
53. Meterko M, Wright S, Lin H, Lowy E, Cleary PD. Mortality among patients with acute myocardial infarction: the influences of patient-centered care and evidence-based medicine. Health Serv Res 2010 Oct;45(5 Pt 1):1188-1204 [FREE Full text] [doi: 10.1111/j.1475-6773.2010.01138.x] [Medline: 20662947]
54. Iversen HH, Bjertnæs OA, Skudal KE. Patient evaluation of hospital outcomes: an analysis of open-ended comments from extreme clusters in a national survey. BMJ Open 2014;4(5):e004848 [FREE Full text] [doi: 10.1136/bmjopen-2014-004848] [Medline: 24879826]
55. Kool R, Kleefstra SM, Borghans I, Atsma F, van de Belt TH. Influence of Intensified Supervision by Health Care Inspectorates on Online Patient Ratings of Hospitals: A Multilevel Study of More Than 43,000 Online Ratings. J Med Internet Res 2016;18(7):e198 [FREE Full text] [doi: 10.2196/jmir.5884] [Medline: 27421302]


## Abbreviations

**IGZ:** The Dutch health care inspectorate
**NPCF:** Federation of Patient and Consumer Organizations



XSL•FO
RenderX

*Edited by G Eysenbach; submitted 22.01.16; peer-reviewed by S Bidmon, K Flott; comments to author 03.04.16; revised version received 20.05.16; accepted 21.06.16; published 20.07.16.*

*Please cite as:*
*Kleefstra SM, Zandbelt LC, Borghans I, de Haes HJ, Kool RB*
*Investigating the Potential Contribution of Patient Rating Sites to Hospital Supervision: Exploratory Results From an Interview Study in the Netherlands*
*J Med Internet Res 2016;18(7):e201*
*URL: http://www.jmir.org/2016/7/e201/*
*doi:10.2196/jmir.5552*
*PMID:27439392*

©Sophia Martine Kleefstra, Linda C. Zandbelt, Ine Borghans, Hanneke J.C.J.M. de Haes, Rudolf B. Kool. Originally published in the Journal of Medical Internet Research (http://www.jmir.org), 20.07.2016. This is an open-access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/2.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on http://www.jmir.org/, as well as this copyright and license information must be included.



XSL•FO
RenderX

Publisher:
JMIR Publications
130 Queens Quay East.
Toronto, ON, M5A 3Y5
Phone: (+1) 416-583-2040
Email: support@jmir.org



XSL•FO
RenderX