Journal of Medical Internet Research

Journal Impact Factor (JIF) (2023): 5.8
Volume 9 (2007), Issue 1 ISSN 1438-8871 Editor in Chief: Gunther Eysenbach, MD, MPH

Contents

Guest Editorial

A Good Death (e6)
David Gustafson. . . . 2

Original Papers

Rates and Determinants of Repeated Participation in a Web-Based Behavior Change Program for Healthy Body Weight and Healthy Lifestyle (e1)
Marieke Verheijden, Marielle Jans, Vincent Hildebrandt, Marijke Hopman-Rock. . . . 6

A Website to Improve Asthma Care by Suggesting Patient Questions for Physicians: Qualitative Analysis of User Experiences (e3)
Christine Hartmann, Christopher Sciamanna, Danielle Blanch, Sarah Mui, Heather Lawless, Michael Manocchia, Rochelle Rosen, Anthony Pietropaoli. . . . 14

Term Identification Methods for Consumer Health Vocabulary Development (e4)
Qing Zeng, Tony Tse, Guy Divita, Alla Keselman, Jon Crowell, Allen Browne, Sergey Goryachev, Long Ngo. . . . 27

Psychological Assessment via the Internet: A Reliability and Validity Study of Online (vs Paper-and-Pencil) Versions of the General Health Questionnaire-28 (GHQ-28) and the Symptoms Check-List-90-Revised (SCL-90-R) (e2)
Miguel Vallejo, Carlos Jordán, Marta Díaz, María Comeche, José Ortega. . . . 36

Assessing Consumer Health Vocabulary Familiarity: An Exploratory Study (e5)
Alla Keselman, Tony Tse, Jon Crowell, Allen Browne, Long Ngo, Qing Zeng. . . . 46

XSL•FO
RenderX

Guest Editorial

A Good Death

David H Gustafson[1], PhD

Center for Health Enhancement Systems Studies, Madison, WI, USA

Corresponding Author:
David H Gustafson, PhD
1513 University Avenue
4105 Mechanical Engineering Building
Madison, WI 53706
USA
Phone: +1-608-2623768
Fax: +1-608-262-8454
Email: dhgustaf@facstaff.wisc.edu

Abstract

The Institute of Medicine defines a good death a “one that is free from avoidable death and suffering for patients, families and caregivers in general accordance with the patients’ and families’ wishes.”. The current system creates barriers to reducing the stress and suffering that accompany a patient’s end of life. Data and eHealth technology, if it were more accessible, could help patients, families, and caregivers to cope with end of life issues.

(J Med Internet Res 2007;9(1):e6) doi:10.2196/jmir.9.1.e6

KEYWORDS

Death; Alzheimer’s Disease; Hospice; Technology; End-of-Life Care; Intensive Care

Introduction

Mom had Alzheimer’s. She was barely hanging on to her apartment in an assisted living facility and would soon have to move to an Alzheimer’s unit. Then breathing troubles started. In three days she was diagnosed with pneumonia, developed congestive heart failure, had a heart attack (one day after being admitted to the hospital), and died. In those three days more money was spent “caring for her” than had been during her entire life.

My dad died when I (the oldest) was 11. Mom raised three kids on a secretary’s salary, put three kids through college, and did so with jokes, smiles, and songs. Her view of life was: make lemonade out of lemons; get on with it; don’t complain.

She took the same approach to death. She got sick on Saturday, spent Sunday and Monday in an Intensive Care Unit, and died Tuesday noon. Even 30 minutes before she died she was trying to make it easy on us by joking. Then she waited to die until we went for to lunch because (I think) she did not want us to see her die. Like always, she did her job and moved on. She was 87. I loved her very much.

A good death?

The Institute of Medicine defined a good death a “one that is free from avoidable suffering for patients, families and caregivers in general accordance with the patients’ and families’ wishes” [1]. I wish I could say my mother got the care she needed and deserved. She did not; neither did we. Those whom I told our experiences suggest they are all-too-common. Hence, I would like to share these experiences here, and discuss the needs, but also responsibilities, of patients and family members at end of a life, and the implications for the health field.

She was lying in that bed. When she first arrived at the hospital, she was still able to walk. I know that even at my age, I can’t go as long without physical exercise as I used to. When I take time off and get back to it, my muscles are sore. How long could she stay in that bed before she would no longer have the muscles to walk out? Would she not be able to return to her assisted living apartment? What about her heart attack? How much would that limit her?

Even in her debilitated mental state, she was probably the most rational person of us all. Mom knew what she wanted: to have the restraints and tubes *removed* and to go home. We were told that we could have none of that. The pneumonia had to be treated. The restraints had to stay on. The tubes had to stay in. Could her antibiotics be given by intramuscular or by pill? What if *we* took out the tubes, stood at the door and told the nurses to leave our mother alone? What would they have done? What rights do family members have and how could we exercise them?

The thing that bothered me the most was those restraints. I wonder if my mother would have had the heart attack if she had not been placed in restraints. Did the hospital “kill” her? Did

XSL•FO
RenderX

the heart attack kill her? Or did she die of a broken heart? Let me explain. I was told that when Mom got to the hospital, she kept pleading to go home. Over and over and over again. The Alzheimer's had made it difficult for her to deal with change and when the nurses added the IVs, she became combative. A team of nurses had to use restraints to subdue a frail 87-year-old woman! Straps held down her hands so she would not pull out the IVs. Mom hated lying on her back, but the restraints made it impossible for her to lie on her side or even to scratch her nose. She continued to fight the restraints for several hours, all the time asking why she was in the hospital and why she could not go home. After a while, she was just exhausted, and shortly after that, she had a heart attack. If I were 87, begging to go home, and doing everything I could do to fight restraints, how long would my heart hold out? Did this have to happen? What rights did the family have? How could we have intervened? Looking back, I wonder what would happen if everyone working in an ICU were required to lie in restraints for just one hour.

I found that not all "do not resuscitate" (DNR) orders are the same. A nurse in the ICU told us that Mom's orders were signed so long ago (seven years ago) that they were not valid. But now she had Alzheimer's and, because of her limited cognitive functioning, had given up power of attorney for health care. Was the nurse right? How were we supposed to know that? What can we do now?

The ICU nurses said that my mother's assisted living facility could not handle a person as sick as she. Again, we had so many unanswered questions. Is it illegal for an assisted living facility to provide such care, even on a temporary basis? Could our family have hired 24-hour nursing care and kept her in the assisted living facility? Where could we find good caregivers even if we were allowed to? How would we know a good home care nurse from a bad one? How would we monitor the care to ensure it was of high quality? What could we have done if poor care was provided?

The nurses told us we would have to raise our questions with the doctors. Yet they seemed to be avoiding us. We called the lead physician's office many times and received no response. We even arrived very early (before 7:00 a.m.) to catch that physician on his rounds; he had "just left." The few communications we did have with doctors were inconsistent and conflicting. The pulmonologist said Mom had 24–48 hours left. The nurses said she was looking better. An internist said she would be home in two days and a psychiatrist just appeared regularly to yell (literally): "Olive, do you know where you are? Can you spell your name?" I interrogated the cardiology nurse. She would not give a prognosis, so I asked "Was there tissue damage?" She said "Yes." I said "Was it a minor heart attack?" She said "No." I said "Was the damage was pretty significant?" She said "Yes." I looked at Mom and thought that she did not look good at all. While she might have been more awake on the second day, she was fighting for air; her chest and abdomen going up and down. I could not see how she was getting better. How could I get a straight answer [2]? How could I take definitive action when I received conflicting advice and my own common sense told me that I was not getting a straight story? How could I get the doctors and nurses to talk to me? Who could I call if they don't? Who was in charge?

It was difficult for us to know whether our mother's condition was due to her disease or her treatment. Mom was clearly out of it. But she was on Haldol, which she had waited several hours to get while the staff attended to another emergency and while she was fighting her restraints. The nursing staff reported that the Haldol helped Mom quiet down. But when I first saw her, her speech was slurred and she looked awful. How could I know whether it was the Haldol or the heart attack or the lack of oxygen to the brain caused by pneumonia? No one would talk with us!

I kept asking myself: "What does Mom want?" Does she want to die? Mom was awake. I could have asked her. But how? Do I say: "Mom, do you want to die?" Or "Have you had enough?" I wanted to approach this the right way, but how do I open the discussion? AND, suppose she had said "Yes!" Then what? What could we have done to help her along?

Was it time to talk to hospice? The community offered several hospice options. Which one should we talk to? How could we find the best one? Was she eligible to go? What would they do for her? If hospice was available, it might be another difficult transition for Mom. I just wanted to ease her misery.

Training medical professionals on the dying process

The death and dying process, including the needs of family members, should have a significant place in the training of physicians and nurses. But when I asked the nurse in charge of the ICU whether the hospital had a palliative care program, she replied: "What is that?" After we complained several times that we could not reach the physician, a "case manager" appeared, asking: "What do you want?" But we did not know what we wanted! She gave a list with telephone numbers of nursing homes and hospices but would not identify the good ones. When we asked whether the good ones were full and all the bad ones empty, she avoided the question. Again, there were so many unanswered questions. How could we get Mom into a nursing home or hospice that would be nice to her? How would we know if they were giving good care? How could we intervene if they weren't? Time was getting short. We needed answers.

I wish that I could be optimistic that things will be better. I am not. The health field has recommended an array of end-of-life policies and best practices [3-6]. The Institute of Medicine and major provider associations, have issued reports, studies, and calls for change. The Robert Wood Johnson Foundation invested millions of dollars to promote a good death. But little has changed. The health care field is full of studies showing that patients' wishes are not respected [7], that communication in the ICU is sorely lacking [8]; that evidence-based practices are not implemented [9]; and that fundamental concepts of palliative care (and even decency) are absent in many health care organizations [10].

Providers (and consumers) need to be more comfortable with uncertainty and with death. Long ago, Larry Weed (developer of the problem oriented medical record and Problem Knowledge Couplers) warned that it was not practical to expect a doctor to store and effectively process the vast and rapidly growing base

XSL•FO
RenderX

of medical knowledge. In many cases, health care providers just don't know how bad things are or the best course of action for a patient. So it is reasonable for a doctor to express uncertainty about what to do or whether the situation is dire enough to call the family together. But many providers act as if death is a failure, when it is part of life [11]. All patients and families deserve honest and consistent information, even if the information is an expression of uncertainty.

Healthcare providers are good people. I have seen my sister, a critical care nurse, come home from her job at a leading teaching hospital so tired she can barely move. I have seen doctors so frustrated with the health system that they can barely see straight. Providers are trained to cure at all costs but the incentives are to ensure that they spend as little time a possible with individual patients. I have seen hospital administrators after another long night of worrying on how they will make budget. If I were in their shoes with the same pressures and incentives, I would probably do the same thing. It should be different but it won't; not for a long time. The problem is the system, not the people.

In the final analysis it is up to us (the families) to take more responsibility for the dying person and for ourselves [12]. We are the ones who care the most about what happens. We are the ones that will make the time. We need to know what to look for and what to expect. We need to know how to care for a patient at the end of life. We need to know our rights and how to exercise them; our options and how to choose between them. We need to know how to assess quality and how to act on those assessments. And we need to understand that death is part of life and that uncertainty is part of the dying process [13].

Ideally, families should be prepared to deal with death and dying well before the event. But it doesn't work that way. I have done health services research for over 30 years. I currently have two research grants on death and dying. I knew my mother was approaching the end of her life. I should have been prepared for her death. I wasn't. I was powerless when it came to making things happen in the ICU. I knew the principles but not the specifics of how to interact with a dying patient, and I needed the specifics. Our family (those in the ICU and those far away) needed ready access to information on Mom's status as well as easy-to-find, easy-to-apply, just-in-time training on death and dying; training that was accessible while we sat in the intensive care unit and included the specific signs to look for and specific words to use with patients and providers.

Recommendations

I have three suggestions: 1) automate the processes for helping patients and families deal with dying and death, 2) when automation is out of the question, make it hard for ICUs to do the wrong thing, and 3) improve transparency of how healthcare deals with death and dying.

Automate

Given the pressures that health care providers operate under, it is unrealistic to expect training and exhortation to change anything. Technologies are needed that equip patients and families to deal with death and dying. Things would have been so much better if I could have opened my smart phone and pulled up a list of the 20 things to watch for in the ICU: things like restraints, conflicting information, care contrary to patient wishes, goals for end of life, family-physician communication. If I could have selected one of those things on my smart phone and seen an overview of the issues, my rights, specific steps I could take. If I could clicked on a topic and received more detail through decision aids, scripts, assessments, and training on how to exercise my rights, all presented in way that I could apply on the spot. If I could have had immediate Web access to databases on ICU, hospice and nursing home quality and relevant literature. If I could have sent a text message to an expert in death and dying or seen video clips of an effective encounter with a patient, a family member, a nurse, doctor or administrator. If I could have asked a question of other families who had gone through something similar or read stories of their experiences. The reality is that I *could have*—the technology, the knowledge, and the data exist to deal with most of these issues.

My wife and I just completed our Health Care Power of Attorney documents and sat down with the kids to review them. I wish that I could have handed the kids a memory stick containing that information, reviewed the basic structure with them, and asked them to carry it with them, because some day (5, 10, 30 years from now—or maybe tomorrow) they will need it, and so will I. Research and development could make this kind of tool a reality. But we can't stop with development. We need to have a system for dissemination. Partnerships with the legal profession could ensure that when a will is prepared or updated that the participants and their families be given access to these tools. Hospitals could make this technology available to families whenever a patient is admitted to intensive care.

Secondly, electronic medical records (EMRs) could improve end of life. Is it out of the question to give families access to the medical record during this difficult time? Could EMRs place the patient's goals for end of life care in an easy to access location that would be hard for any provider to miss? Could EMR's have reminder systems that would alert providers when it is time to encourage patients to update their end of life wishes? Whenever a patient is considered to have a potentially life-threatening condition, could the EMR require documentation from ICU clinicians indicating that they have read and accept the patient's goals for care at end of life?

Make it hard to do the wrong thing

Systems change can be difficult. However, there are principles that can increase the likelihood of success. One is to remove the status quo. For instance, one cannot use mechanical restraints if they are not available. Administration could remove the mechanical restraints from the ICU and lock them in a cabinet for use only with permission of senior leaders. Medicare could treat inappropriate care at the end of life in the same way they treat other medical errors. Medicare policy could just flat out prohibit the use of mechanical restraints without permission from a senior leader of the hospital. Then, immediate convening of a rapid response team could be required to determine steps needed to remove the restraints and make sure they are never needed again. These steps have already been taken by mental

XSL•FO
RenderX

health hospitals; we could do the same for dying patients. Technology could help to ensure that these policies are implemented.

Transparency

Public reporting on the quality of a death should be required. Organizations like Medicare and NCQA could collect and publicly report data on use of restraints and other measurable dimensions of quality in death and dying. But public reporting is just a start. Systems must be in place to ensure that people will actually act on this information. Given our reticence to address death and dying before it happens, it is unrealistic to expect families to study quality of death data until the time comes. Hence, it will be important to find ways to make these data and resources easily available, easily understandable and easy to act on in a just-in-time basis.

We (the patients and families) need to take responsibility for our own dying and death. It is the centerpiece for us being able to do that. Mom needed it, I needed it and so will you.

References

1. Field M, Cassell C. Approaching death: improving care at the end of life. (IOM Report) Washington DC: National Academy Press; 1997.
2. Kirchhoff KT, Spuhler V, Walker L, Hutton A, Cole BV, Clemmer T. Intensive care nurses experiences with end of life care. Am J Crit Care 2000;9:36-42. [Medline: 10631389]
3. Kirchhoff KT. Promoting a peaceful death in the ICU. Crit Care Nurs Clin North Am 2002 Jun;14(2):201-206. [Medline: 12038508] [doi: 10.1016/S0899-5885(01)00012-0]
4. Jackson I. Bereavement follow-up support in intensive care units. Intensive Crit Care Nsg 1992;8(3):163-168. [doi: 10.1016/0964-3397(92)90023-D]
5. ; American Medical Association. Education for physicians on end of life care (EPERC). 1999. URL: http://www.epec.net/EPEC/Webpages/ph.cfm
6. Curtis JR, Patrick DL, Shannon SE, Treece PD, Engelberg RA, Rubenfeld GD. The family conference as a focus to improve communication about end-of-life care in the intensive care unit: opportunities for improvement. Crit Care Med 2001 Feb;29(2 Suppl) N26-33. [Medline: 11228570] [doi: 10.1097/00003246-200102001-00006]
7. Lynn J, Teno JM, Phillips RS, Wu AW, Desbiens N, Harrold J, et al. Perceptions by family members of the dying experience of older and seriously ill patients. SUPPORT Investigators. Study to Understand Prognoses and Preferences for Outcomes and Risks of Treatments. Ann Intern Med 1997 Jan 15;126(2):97-106 [FREE Full text] [Medline: 9005760]
8. Kirchhoff KT, Walker L, Hutton A, Spuhler V, Cole BV, Clemmer T. The vortex: families' experiences with death in the intensive care unit. Am J Crit Care 2002 May;11(3):200-209. [Medline: 12022483]
9. McGlynn EA, Asch SM, Adams J, Keesey J, Hicks J, DeCristofaro A, et al. The quality of health care delivered to adults in the United States. New Engl J Med 2003;26:2635-2645. [Medline: 12826639]
10. Jacobson JA, Francis LP, Battin MP, Green GJ, Grammes C, Vanriper J, et al. Dialogue to action: lessons learned from some family members of deceased patients at an interactive program in seven Utah hospitals. J Clin Ethics 1997;8(4):359-371. [Medline: 9503086]
11. Asch DA, Hansen-Flaschen J, Lanken PN. Decisions to limit or continue life-sustaining treatment by critical care physicians in the United States: conflicts between physicians' practices and patients' wishes. Am J Respir Crit Care Med 1995 Feb;151(2 Pt 1):288-292. [Medline: 7842181]
12. Parker L. In a crisis do-not-revive requests do not always work. USA Today 2006 Dec 20.
13. Gross J. For families of the dying, coaching as the hours wane. New York Times 2006 May 20.

Submitted 01.09.07; this is a non–peer-reviewed article;accepted 26.02.07; published 14.03.07.

Please cite as:
Gustafson DH
A Good Death
J Med Internet Res 2007;9(1):e6
URL: http://www.jmir.org/2007/1/e6/
doi:10.2196/jmir.9.1.e6
PMID:17478415

© David H Gustafson. Originally published in the Journal of Medical Internet Research (http://www.jmir.org), 14.03.2007. Except where otherwise noted, articles published in the Journal of Medical Internet Research are distributed under the terms of the Creative Commons Attribution License (http://www.creativecommons.org/licenses/by/2.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited, including full bibliographic details and the URL (see "please cite as" above), and this statement is included.

XSL•FO
RenderX

Original Paper

Rates and Determinants of Repeated Participation in a Web-Based Behavior Change Program for Healthy Body Weight and Healthy Lifestyle

Marieke W Verheijden[1], PhD; Marielle P Jans[1], PhD; Vincent H Hildebrandt[1,2], PhD, MD; Marijke Hopman-Rock[1,2], PhD

[1]TNO Quality of Life, Leiden, The Netherlands

[2]Body@Work, Research Centre Physical Activity, Work and Health, TNO-VU University Medical Centre, Amsterdam, The Netherlands

Corresponding Author:
Marieke W Verheijden, PhD
TNO Quality of Life
PO Box 2215
2301 CE Leiden
The Netherlands
Phone: + 31 71 518 1706
Fax: +31 71 518 1903
Email: Marieke.Verheijden@tno.nl

Abstract

Background: In recent years, many tailored lifestyle counseling programs have become available through the Internet. Previous research into such programs has shown selective enrollment of relatively healthy people. However, because of the known dose-response relationship between the intensity and frequency of counseling and the behavior change outcomes, selective retention may also be a concern.

Objective: The aim of this study was to identify rates and determinants of repeat participation in a Web-based health behavior change program.

Methods: A Web-based health behavior change program aimed to increase people's awareness of their own lifestyle, to promote physical activity, and to prevent overweight and obesity was available on the Internet from July 2004 onward at no cost. Univariate and multivariate logistic regression analyses were conducted to identify characteristics of people who participated in the program more than once. Age, compliance with physical activity guidelines, body mass index, smoking status, and the consumption of fruit, vegetables, and alcohol were included in the analyses.

Results: A total of 9774 people participated in the baseline test, of which 940 used the site more than once (9.6%). After exclusion of individuals with incomplete data, 6272 persons were included in the analyses. Of these 6272 people, 5560 completed only the baseline test and 712 also participated in follow-up. Logistic regression predicting repeated use determined that older individuals were more likely to participate in follow-up than people aged 15-20 years. The odds ratios for the age categories 41-50, 51-60, and > 60 years were 1.40 (95% CI = 1.02-1.91), 1.43 (95% CI = 1.02-2.01), and 1.68 (95% CI = 1.03-2.72), respectively. Individuals who never smoked were more likely to participate repeatedly than current smokers and ex-smokers (OR = 1.44, 95% CI = 1.14-1.82 and OR = 1.49, 95% CI = 1.17-1.89, respectively). People meeting the guidelines for physical activity of moderate intensity (OR = 1.23 95% CI = 1.04-1.46) and for vegetable consumption (OR = 1.26 95% CI = 1.01-1.57) were also more likely to participate repeatedly than people who did not, as were obese people compared to individuals with normal weight (OR = 1.41 95% CI = 1.09-1.82).

Conclusions: For some variables, this study confirms our concern that behavioral intervention programs may reach those who need them the least. However, contrary to most expectations, we found that obese people were more likely to participate in follow-up than people of normal body weight. The non-stigmatizing way of addressing body weight through the Internet may be part of the explanation for this. Our findings suggest that Web-based health behavior change programs may be more successful in the area of weight management than in many other health-related areas. They also stress the importance of adequate coverage of weight management in Web-based health promotion programs, as a driver to continue participation for overweight and obese people.

(J Med Internet Res 2007;9(1):e1) doi:10.2196/jmir.9.1.e1

XSL•FO
RenderX

KEYWORDS

Internet; counseling; health behavior change; selective enrollment; selective retention

Introduction

In recent years, many tailored health promotion programs have become available through the Internet. As with any other health-promoting intervention, these Web-based health promotion programs are not expected to lead to either short-term or sustained behavior change unless the intervention reaches the intended target audience. Unfortunately, it has been postulated that health-promoting lifestyle interventions tend to reach those who need them the least [1]. For ethical and practical reasons, there is often no information available on people who decide not to participate (see, for example, the study by Sirard and colleagues [2]). Dutta-Bergman [3] showed that people who look for health information on the Internet are more health oriented than people who don't look for such information. They are also more likely to hold stronger health-oriented beliefs and to engage in healthy activities. Similarly, Verheijden et al [4] conducted a nonresponse survey and showed that participants in a Web-based tailored nutrition counseling program were a relatively well-educated and healthy subsample of the target audience. It is thus evident that efforts need to be made to minimize selective enrollment in Web-based health promotion programs.

Furthermore, it is known that there is a strong dose-response relationship between the number and intensity of counseling sessions and behavior change outcomes [5-7]. It is therefore unlikely that a difficult process such as health behavior change can be achieved in a single counseling session; repeated participation is likely necessary to achieve sustainable changes. It seems reasonable to assume that this is also true for Web-based counseling, which means that repeated exposure to the counseling programs is necessary to achieve sustainable changes.

In addition to selective enrollment in Web-based programs, selective attrition during follow-up may thus be a concern too. This concern was also expressed by Eysenbach [8] and by Danaher and colleagues [9], who argued that attrition, uptake, exposure, and diffusion measures need to be addressed in addition to the effectiveness of eHealth programs. Indeed, little is known about attrition and its determinants in Web-based behavior change programs. The current study therefore addresses the rates and health and lifestyle determinants of repeated participation in a Web-based health behavior change program. The current focus on user characteristics is in line with the relevance of these characteristics that was made explicit by Christenson and Mackinnon [10].

Methods

The Web-Based Health Promotion Program

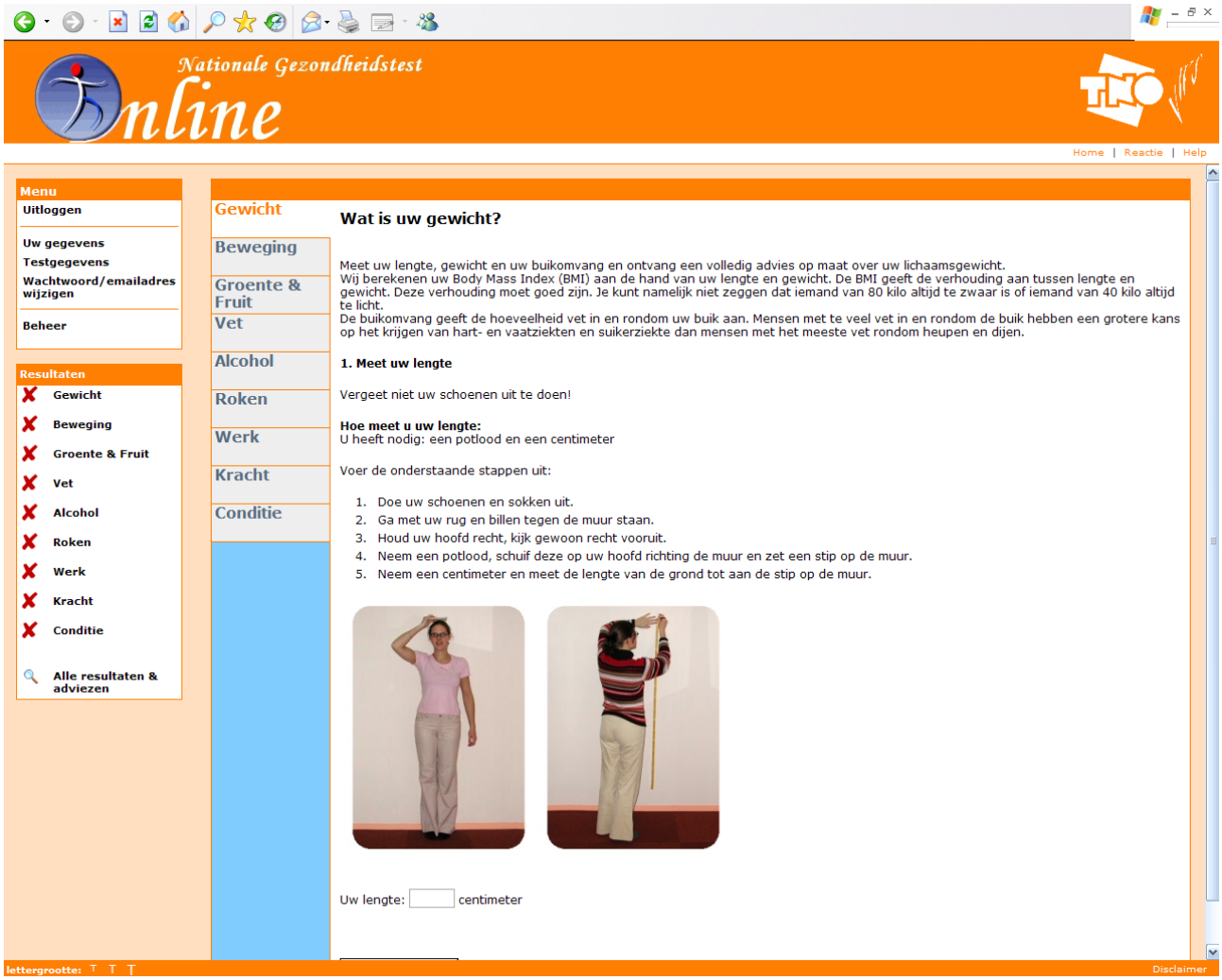
The Web-based health promotion program on which the current paper is based was a Web-based version of the Dutch National Health Test, designed by the Netherlands Organisation for Applied Scientific Research TNO in cooperation with the Dutch Foundation Pur Sang. The program was developed with funding from the Dutch Ministry of Health, Welfare and Sport and aimed to increase people's awareness of their own lifestyle, to promote physical activity, and to prevent overweight and obesity. The program was available on the Internet from July 2004 onward at no cost. The launch of the site was brought to the public's attention with a press release presenting the State Secretary as the very first participant in the program. No further media and marketing strategies were used to keep the program in the public's view. However, various organizations were interested in the program, which generated free publicity. For example, articles on the Dutch National Health Test were published in national and local newspapers, in free magazines published by several supermarket chains, in women's magazines, and on websites of general practices and municipal health services.

The registration procedure for the website included selection of a personal username and password and a series of questions on sociodemographic characteristics of the participants. The health promotion program contained modules on anthropometrics (height, weight, waist circumference), physical activity, dietary habits, alcohol intake, smoking, work, cardiorespiratory fitness, and muscle strength. The modules on lifestyle consisted of a series of questions to assess the current behavior. The modules on anthropometrics, cardiorespiratory fitness, and muscle strength contained instructions for the appropriate self-tests. Upon completion of each individual module, participants received feedback that was tailored to the responses they had given in that module. Because the physical activity module was the core of the program, additional questions on physical activity were included. Figure 1 presents a screenshot of the Dutch National Health Test (in Dutch). Additional screenshots are available in the Multimedia Appendix (in Dutch). The Dutch National Health Test is no longer available to the public. More information can be obtained from the authors.

The tailored feedback messages for the individual modules were also integrated in an overall report. Participants could print this report and keep it for their own reference. The data were stored in a database and used as a basis for longitudinal feedback in follow-up participation. The data were also kept for research purposes. Participants' consent for this was obtained. Participants were allowed to complete the modules over a two-week period. They were sent an email reminder to complete the modules in a timely manner. During the initial participation, people were made aware of the availability of follow-up modules. They were encouraged to participate in these follow-up modules to monitor their progress and to receive more tailored feedback. Invitations for follow-up participation were sent out by email three months after the completion date of the last entry. The program was made available online in June 2004, which means that the first reminders for follow-up could have been sent out in October 2004; however, due to a technical error in the automated reminders, the first reminders were not sent out until March 2005.

XSL•FO
RenderX

Figure 1. Screenshot of the Dutch National Health Test (Nationale Gezondheidstest Online)



The Independent Variable: Participation

Although each individual module led to immediate tailored feedback, people were not considered to be participants unless they had completed the modules on anthropometrics and physical activity. Single participation was therefore defined as having only one record in the Dutch National Health Test database in which the modules on anthropometrics and physical activity were completed. Repeat participation was defined as having two or more records in the Dutch National Health Test in which the modules on anthropometrics and physical activity were completed.

Predictors for Repeat Participation

Data on gender, age (15-20, 21-30, 31-40, 41-50, 51-60, 61 and older), and education level (very low, low, intermediate, high, very high) were obtained in the registration process of the Web-based health promotion program. The body mass index was calculated using data on self-reported height and weight. It was subsequently categorized (≤ 25 kg/m^2, 25.01-30 kg/m^2, > 30 kg/m^2). People were encouraged to measure and report their waist circumference (in centimeters). Detailed instructions on measuring waist circumference, which included some clear pictures, were provided. Smoking status was defined as currently smoking, formerly smoking, or never having smoked. Physical activity was categorized based on the criteria for sufficient physical activity of (1) moderate intensity (moderate intense physical activity at least 5 days per week for at least 30 minutes per day) and (2) high intensity (intense physical activity at least 3 days per week for at least 20 consecutive minutes). Current Dutch guidelines define minimum intake levels for sufficient fruit and vegetables. For fruit, the guideline is a minimum of two pieces per day; for vegetables, the guideline is a minimum of 200 g per day. Participants were categorized as either meeting or not meeting the guidelines for fruit and vegetable consumption. Alcohol consumption was defined based on current Dutch gender-specific guidelines, which define the maximum number of alcoholic drinks per week (ie, a maximum of 15 alcoholic drinks per week for women and 21 for men). People exceeding these numbers were defined as excessive drinkers.

Analysis

Descriptive statistics were used to present data on the baseline characteristics of the participants of the Web-based health promotion program. No data on waist circumference will be presented because inspection of the data revealed that people likely gave invalid answers. For example, the reported waist circumferences varied from as little as 20 cm to as much as 2 km. Furthermore, values such as 50 cm, 60 cm, 70 cm, and 80

XSL•FO
RenderX

cm were reported much more frequently than values such as 53 cm, 69, and 77 cm. Descriptive statistics were also used to present data on single and repeated use of the Web-based health promotion programs. These analyses are based on data from the 9774 people who met the participation requirements of this program (ie, people who completed questions on anthropometrics and physical activity).

Univariate and multivariate logistic regression analyses were conducted to identify the crude and adjusted effects of factors associated with repeat participation in a Web-based health behavior change program. A total of 9774 people enrolled in the Web-based health promotion program, but because of missing data, all regression analyses were based on 6272 participants. The exclusion of 3502 people was the result of missing values in the variables on smoking (missing for 3010 people) and on the consumption of fruit (missing for 2751 people), vegetables (missing for 2751 people), and alcohol (missing for 3399 people). Analyses comparing the characteristics of people with and without missing values revealed no clinically relevant differences. People with missing values, for example, were significantly older than people without missing values ($P < 0.01$), but the difference was less than one year.

Results

Participants and Number of Visits to the Web-Based Health Promotion Program

Approximately two thirds of the 9774 participants were female. The mean age was 36 years (SD = 13). The vast majority of the participants (90.1%) had an intermediate or (very) high education level. The mean body mass index was 24.5 kg/m^2 (SD = 5.7). At the time of enrollment, 22% of the participants were smokers. The guidelines for physical activity of moderate intensity and high intensity were met by 51% and 46% of participants, respectively. Few people met the guidelines for fruit and vegetable consumption (22% and 14%, respectively), and 7% consumed more alcoholic drinks per week than recommended by current Dutch gender-specific guidelines.

Of the 9774 people who enrolled in the Web-based health promotion program, almost 10% participated more than once: 7.6% participated twice, and 1.9% participated three times. The completion of four visits was very infrequent (< 1%).

Determinants of Repeat Participation

People aged 41 years and older repeatedly participated in the intervention more than those aged 15-20 years (Table 1). In the univariate analyses, a healthy lifestyle was related to repeat participation. For example, repeat participation was more frequent among former smokers (OR = 1.73) and people who never smoked (OR = 1.47) than among current smokers. It was also more frequent among people with sufficient physical activity than among people with insufficient physical activity (OR = 1.31 for moderate intensity activity, OR = 1.23 for high intensity activity). Finally, repeated participation was more frequent among people meeting the guidelines for fruit consumption (OR = 1.26) and vegetable consumption (OR = 1.39) than among those failing to meet the guidelines. In contrast to the pattern that a healthy lifestyle was related to repeat participation, people who were overweight (OR = 1.20) or obese (OR = 1.54) more frequently participated repeatedly than people of normal body weight.

In the multivariate analyses, repeat participation was more frequent among people aged 41 years and older (OR = 1.40-1.68), among obese people (OR = 1.41), among former smokers and individuals who never smoked (OR = 1.49 and 1.44, respectively), among people with insufficient physical activity of moderate intensity (OR = 1.23), and among people with a sufficient vegetable consumption (OR = 1.26) than among the relevant reference groups.

XSL•FO
RenderX

Table 1. Participants' baseline characteristics as determinants of repeat participation in a Web-based health behavior change program (bold numbers indicate significant OR values), N = 6272

	Repeat Participation (%)	Crude Effect		Adjusted Effect*	
		OR	95% CI	OR	95% CI
Gender					
Male (n = 2132)	11.7	1.00	–	1.00	–
Female (n = 4140)	11.2	0.95	0.80-1.11	0.99	0.83-1.18
Age (years)					
15-20 (n = 846)	9.3	1.00	–	1.00	–
21-30 (n = 1717)	8.6	0.92	0.69-1.22	0.97	0.72-1.32
31-40 (n = 1449)	12.0	1.33	1.00-1.75	1.31	0.97-1.79
41-50 (n = 1281)	13.1	**1.47**	**1.11-1.95**	**1.40**	**1.02-1.91**
51-60 (n = 791)	14.2	**1.60**	**1.18-2.18**	**1.43**	**1.02-2.01**
61 and older (n = 188)	16.6	**1.92**	**1.23-3.03**	**1.68**	**1.03-2.72**
Education level					
Very low (n = 159)	12.6	1.00	–	1.00	–
Low (n = 421)	12.6	1.00	0.58-1.74	0.90	0.51-1.58
Intermediate (n = 1838)	11.8	0.93	0.57-1.52	0.88	0.53-1.46
High (n = 1082)	10.8	0.84	0.51-1.40	0.83	0.49-1.39
Very high (n = 2272)	11.0	0.86	0.53-1.39	0.77	0.46-1.27
Body mass index (kg/m^2)					
≤ 25 (n = 4044)	10.4	1.00	–	1.00	–
25.01-30 (n = 1662)	12.3	**1.20**	**1.01-1.44**	1.11	0.92-1.33
> 30 (n = 566)	15.2	**1.54**	**1.20-1.88**	**1.41**	**1.09-1.82**
Smoking status					
Current smoker (n = 1343)	8.1	1.00	–	1.00	–
Former smoker (n = 2092)	13.2	**1.73**	**1.37-2.18**	**1.49**	**1.17-1.89**
Never smoker (n = 2837)	11.5	**1.47**	**1.17-1.84**	**1.44**	**1.14-1.82**
Physical activity, moderate intensity†					
Insufficient (n = 3009)	9.9	1.00	–	1.00	–-
Sufficient (n = 3263)	12.7	**1.31**	**1.12-1.54**	**1.23**	**1.04-1.46**
Physical activity, high intensity‡					
Insufficient (n = 3377)	10.4	1.00	–	1.00	–
Sufficient (n = 2895)	12.5	**1.23**	**1.05-1.44**	1.11	0.94-1.31
Vegetable consumption§					
Insufficient (n = 5409)	10.9	1.00	–	1.00	–
Sufficient (n = 863)	14.5	**1.39**	**1.13-1.71**	**1.26**	**1.01-1.57**
Fruit consumption‖					
Insufficient (n = 4919)	10.8	1.00	–	1.00	–
Sufficient (n = 1353)	13.2	**1.26**	**1.05-1.50**	1.06	0.87-1.28
Alcohol consumption¶					
Excessive (n = 420)	11.2	1.00	–	1.00	–

XSL•FO
RenderX

	Repeat Participation (%)	Crude Effect		Adjusted Effect*	
		OR	95% CI	OR	95% CI
Moderate or none (n = 5852)	11.4	0.98	0.72-1.35	1.02	0.74-1.41

*The regression model contained repeat participation (yes/no) as the dependent variable and gender, age, education level, body mass index, smoking status, physical activity (moderate and high intensity), fruit consumption, vegetable consumption, and alcohol consumption as independent variables.

†Current guidelines in The Netherlands recommend a total of at least 30 minutes of physical activity of moderate intensity at least 5 days per week.

‡Current guidelines in The Netherlands recommend at least 20 consecutive minutes of physical activity of high intensity at least 3 days per week.

§Current guidelines in The Netherlands recommend at least 200 g of vegetables per day.

‖Current guidelines in The Netherlands recommend at least 2 pieces of fruit per day.

¶Current guidelines in The Netherlands recommend a maximum of 15 alcoholic drinks per week for women and 21 drinks per week for men.

Discussion

This study showed that people who repeatedly participated in a Web-based health promotion program generally had healthier lifestyles than people who participated only once. In contrast to this and to our expectations, people who were overweight or obese participated more frequently than people of normal body weight. Repeated use was relatively infrequent; approximately 10% of the people used the program more than once.

The initial concern in reaching the appropriate target audience with Web-based health promotion programs is to prevent selective enrollment in the program. An extensive comparison of the baseline characteristics of the participants in the current study with the Dutch population in general is beyond the scope of this article. A birds-eye view of the baseline characteristics, however, indicates that the participants in the Web-based program had a lower prevalence of overweight and obesity and smoking, and a higher prevalence of compliance with the guidelines for physical activity and fruit and vegetable consumption than the Dutch population in general.

As was discussed recently by Eysenbach [8], eHealth applications face the difficulty that a (sometimes substantial) proportion of people will not be using the application or will be using it sparingly. The latter was also true for the Dutch National Health Test. Our study confirms our hypothesis that selective retention in Web-based health behavior change programs is a concern in addition to selective enrollment. The group of participants who used the program repeatedly were a relatively healthy subsample of all people who enrolled in the program. The only exception to this was for body weight, as people who were overweight or obese used the program more frequently than people of normal body weight. One explanation for this is that a higher risk for disease is associated with a higher probability of participating in counseling [11]. It is known that overweight and obese people perceiving weight as a health risk are more likely to have prepared and/or initiated activities to lose weight [12]. Furthermore, people with chronic conditions are more likely to search for health information on the Internet than those without [13]. Another possible explanation for the fact that overweight and obese people used the program relatively frequently is that Web-based counseling may be particularly appealing for people with stigmatizing diseases. Despite the increasing prevalence of overweight and obesity in most groups of the population [14,15], excessive body weight or the failure to lose weight may continue to be stigmatizing [16-18]. This may also help explain the unexpected effects that were observed for body weight in the current study.

Given the known dose-response relationship between the frequency and intensity of counseling and the achieved behavior change outcomes, it is disappointing that only 10% of the participants used the program more than once. Previous research on Web-based health behavior change programs has shown that people are much less interested in programs that encourage lifestyle improvement than they are in programs that simply compare their behavior to relevant guidelines [19]. This comparison to relevant guidelines can be achieved with a single participation. On the other hand, when people are looking for solid counseling on possible lifestyle improvements, multiple counseling sessions are necessary. People's lack of interest in behavior change counseling may thus help to explain the limited repeated use of the current program.

It is unclear how people's motivation to participate in behavior change counseling may be increased, but it is evident that this change needs to be brought about before Web-based health promotion programs have the potential to lead to sustained behavior change. Upon first use of the program, an effort should be made to explain that behavior change does not occur overnight and that the program people are working with includes follow-up modules that make long-term support possible. Work presented by Spittaels and De Bourdeaudhuij [20] suggests some other approaches that may contribute to increased use of Web-based behavior change programs. A key issue may be to have face-to-face contact before people are referred to the Web-based program. When 100 flyers with information on a Web-based program to promote physical activity were handed out to participants in person, it led to 41 people receiving tailored advice. When the same number of flyers were placed in strategic positions throughout a hospital, it led to only 8 people receiving tailored advice. Another factor that was emphasized in the work of Spittaels and De Bourdeaudhuij [20] is the use of frequent reminder emails. These reminder emails were appreciated by the study participants, but no effect in terms of self-reported physical activity was observed. Rewarding people with something may also help to increase repeated participation [19]. Verheijden and colleagues reported that 76% of the people who were not intrinsically motivated to participate in follow-up programs said they would be interested in participating when given a bonus or reward.

XSL•FO
RenderX

In conclusion, our findings support the debate on the current proliferation of Web-based health behavior change programs and they stress the need to find new approaches to reach the primary target groups via the Web. This study supports earlier findings that Web-based health behavior change programs may largely fail to reach those for whom health behavior change is most necessary. By interesting contrast, overweight and obese people were more frequently repeat users than people of normal body weight. This effect may be due in part to the non-stigmatizing nature of Web-based interventions as opposed to face-to-face interventions. These findings suggest that Web-based health behavior change programs may be more successful in the area of weight management than in many other health-related areas. It also stresses the importance of adequate coverage of weight management in Web-based health promotion programs, as a driver to continue participation for overweight and obese people.

Acknowledgments

The authors gratefully acknowledge the financial support of the Dutch Ministry of Health, Welfare and Sport. The authors also acknowledge Gert Jacobusse for his statistical advice and Richard van den Winkel for his help preparing Figure 1 and the Multimedia Appendix.

Conflicts of Interest

None declared.

Multimedia Appendix

Screenshots of the Dutch National Health Test [PPT file (MS PowerPoint), 848 KB - jmir_v9i1e1_app1.ppt]

References

1. Koelen MA, van den Ban AW. Health education and health promotion. Wageningen, The Netherlands: Wageningen Academic Publishers; 2004.
2. Sirard JR, Pfeiffer KA, Pate RR. Motivational factors associated with sports program participation in middle school students. J Adolesc Health 2006 Jun;38(6):696-703. [Medline: 16730598] [doi: 10.1016/j.jadohealth.2005.07.013]
3. Dutta-Bergman MJ. Health attitudes, health cognitions, and health behaviors among Internet health information seekers: population-based survey. J Med Internet Res 2004 May 28;6(2):e15 [FREE Full text] [Medline: 15249264] [doi: 10.2196/jmir.6.2.e15]
4. Verheijden M, Bakx JC, Akkermans R, van den Hoogen H, Godwin NM, Rosser W, et al. Web-based targeted nutrition counselling and social support for patients at increased cardiovascular risk in general practice: randomized controlled trial. J Med Internet Res 2004 Dec 16;6(4):e44 [FREE Full text] [Medline: 15631968] [doi: 10.2196/jmir.6.4.e44]
5. Draper M, Jennings J, Barón A, Erdur O, Shankar L. Dose-effect relationships in brief therapy based on a nationwide college counseling center sample. Austin, TX: The Counseling & Mental Health Center, The University of Texas at Austin; 2000.
6. Holtrop JS, Wadland WC, Vansen S, Weismantel D, Fadel H. Recruiting health plan members receiving pharmacotherapy into smoking cessation counseling. Am J Manag Care 2005 Aug;11(8):501-507 [FREE Full text] [Medline: 16095436]
7. Rodondi N, Humair JP, Ghali WA, Ruffieux C, Stoianov R, Seematter-Bagnoud L, et al. Counselling overweight and obese patients in primary care: a prospective cohort study. Eur J Cardiovasc Prev Rehabil 2006 Apr;13(2):222-228. [Medline: 16575276] [doi: 10.1097/01.hjr.0000209819.13196.a4]
8. Eysenbach G. The law of attrition. J Med Internet Res 2005;7(1):e11 [FREE Full text] [Medline: 15829473] [doi: 10.2196/jmir.7.1.e11]
9. Danaher BG, Boles SM, Akers L, Gordon JS, Severson HH. Defining participant exposure measures in Web-based health behavior change programs. J Med Internet Res 2006;8(3):e15 [FREE Full text] [Medline: 16954125] [doi: 10.2196/jmir.8.3.e15]
10. Christensen H, MacKinnon A. The law of attrition revisited. J Med Internet Res 2006;8(3):e20 ; author reply e21 [FREE Full text] [Medline: 17032636] [doi: 10.2196/jmir.8.3.e20]
11. Hughes Halbert C, Kessler LJ, Collier A, Brewster K, Weathers B. Factors associated with participation in cancer genetic counselling among African American women. Proceedings from the Annual Clinical Genetics Meeting of the American College of Medical Genetics,. Dallas, Texas; 2005 Mar 17. URL: http://www.acmg.net/resources/ACMG/2005/abs-group-05.pdf [accessed 2007 Jan 6]
12. Wee CC, Davis RB, Phillips RS. Stage of readiness to control weight and adopt weight control behaviors in primary care. J Gen Intern Med 2005 May;20(5):410-415 [FREE Full text] [Medline: 15963162] [doi: 10.1111/j.1525-1497.2005.0074.x]
13. 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. [Medline: 16704514] [doi: 10.1111/j.1475-6773.2006.00510.x]

XSL•FO
RenderX

14. Torrance GM, Hooper MD, Reeder BA. Trends in overweight and obesity among adults in Canada (1970-1992): evidence from national surveys using measured height and weight. Int J Obes Relat Metab Disord 2002 Jun;26(6):797-804 [FREE Full text] [Medline: 12037650] [doi: 10.1038/sj.ijo.0801991]
15. Ogden CL, Carroll MD, Curtin LR, McDowell MA, Tabak CJ, Flegal KM. Prevalence of overweight and obesity in the United States, 1999-2004. JAMA 2006 Apr 5;295(13):1549-1555. [Medline: 16595758] [doi: 10.1001/jama.295.13.1549]
16. Carr D, Friedman MA. Is obesity stigmatizing? Body weight, perceived discrimination, and psychological well-being in the United States. J Health Soc Behav 2005 Sep;46(3):244-259. [Medline: 16259147]
17. Vaidya V. Psychosocial aspects of obesity. Adv Psychosom Med 2006;27:73-85. [Medline: 16418544] [doi: 10.1159/000090965]
18. Griffiths F, Lindenmeyer A, Powell J, Lowe P, Thorogood M. Why are health care interventions delivered over the internet? A systematic review of the published literature. J Med Internet Res 2006;8(2):e10 [FREE Full text] [Medline: 16867965] [doi: 10.2196/jmir.8.2.e10]
19. Verheijden MW, Jans MP, Hildebrandt VH. Web-Based Tailored Lifestyle Programs: Exploration of the Target Group's Interests and Implications for Practice. Health Promot Pract 2006 Jun 30. [Medline: 16816030]
20. Spittaels H, De Bourdeaudhuij I. Volwassenen tot beweging stimuleren via het Internet. In: Steens G, editor. Moet er nog sport zijn? Sport, beweging en gezondheid in Vlaanderen. Volume 1. Antwerp, Belgium: F&G Partners; 2006.

submitted 26.06.06; peer-reviewed by A Steptoe; comments to author 07.08.06; revised version received 13.11.06; accepted 20.12.06; published 22.01.07.

Please cite as:
Verheijden MW, Jans MP, Hildebrandt VH, Hopman-Rock M
Rates and Determinants of Repeated Participation in a Web-Based Behavior Change Program for Healthy Body Weight and Healthy Lifestyle
J Med Internet Res 2007;9(1):e1
URL: http://www.jmir.org/2007/1/e1/
doi:10.2196/jmir.9.1.e1
PMID:

© Marieke W Verheijden, Marielle P Jans, Vincent H Hildebrandt, Marijke Hopman-Rock. Originally published in the Journal of Medical Internet Research (http://www.jmir.org), 22.01.07. Except where otherwise noted, articles published in the Journal of Medical Internet Research are distributed under the terms of the Creative Commons Attribution License (http://www.creativecommons.org/licenses/by/2.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited, including full bibliographic details and the URL (see "please cite as" above), and this statement is included.

XSL•FO
RenderX

Original Paper

A Website to Improve Asthma Care by Suggesting Patient Questions for Physicians: Qualitative Analysis of User Experiences

Christine W Hartmann[1], MSS, PhD; Christopher N Sciamanna[1], MD, MPH; Danielle C Blanch[2], BA; Sarah Mui[3], BA; Heather Lawless[4], BA; Michael Manocchia[5], PhD; Rochelle K Rosen[6], PhD; Anthony Pietropaoli[7], MD

[1]Jefferson Medical College, Philadelphia, PA, USA
[2]Brigham and Women's Hospital, Boston, MA, USA
[3]University of Massachusetts Medical School, Worcester, MA, USA
[4]Wheaton College, Norton, MA, USA
[5]Health Dialog Analytic Solutions, Boston, MA, USA
[6]The Miriam Hospital, Providence, RI, USA
[7]University of Rochester Medical Center, Rochester, NY, USA

Corresponding Author:
Christine W Hartmann, MSS, PhD
Department of Health Policy
Jefferson Medical College
1015 Walnut Street, Suite 115
Philadelphia, PA 19107
USA
Phone: +1 215 955 5744
Fax: +1 215 923 7583
Email: cwhrtmnn@aol.com

Abstract

Background: Asthma is one of the most prevalent chronic conditions in the United Sates, yet despite the existence of national guidelines, nearly three fourths of patients with asthma do not have adequate control and clinical adherence to guidelines is low. While there are many reasons for this, physician inertia with respect to treatment change is partly to blame. Research suggests that patients who ask for specific tests and treatments are more likely to receive them.

Objectives: This study investigated the impact and experience of using an interactive patient website designed to give patients individual feedback about their condition and to suggest tailored questions for patients to ask their physician. The website was designed to be used prior to a physician visit, to increase the likelihood that patients would receive recommended tests and treatments.

Methods: A total of 37 adult patients with asthma participated in semi-structured telephone interviews aimed at eliciting information about their experiences with the website. Transcripts were coded using qualitative data analysis techniques and software. Themes were developed from subsets of codes generated through the analysis. In addition, 26 physicians were surveyed regarding their impressions of the website.

Results: Opportunities exist for improving website feedback, although the majority of both patient and physician respondents held favorable opinions about the site. Two major themes emerged regarding patients' experiences with the website. First, many patients who used the website had a positive shift in their attitudes regarding interactions with their physicians. Second, use of the website prompted patients to become more actively involved in their asthma care. No patient reported any negative experiences as a result of using the website. Physicians rated the website positively.

Conclusions: Patients perceived that the interactive website intervention improved communication and interaction with their physicians, suggesting that patients can play a role in overcoming the clinical inertia of providers. Although the design and content of the website can be improved upon, the main findings suggest that use of the website is well accepted and is perceived to improve the quality of care that patients receive.

(J Med Internet Res 2007;9(1):e3) doi:10.2196/jmir.9.1.e3

XSL•FO
RenderX

KEYWORDS

Asthma; Internet; qualitative research; patient education

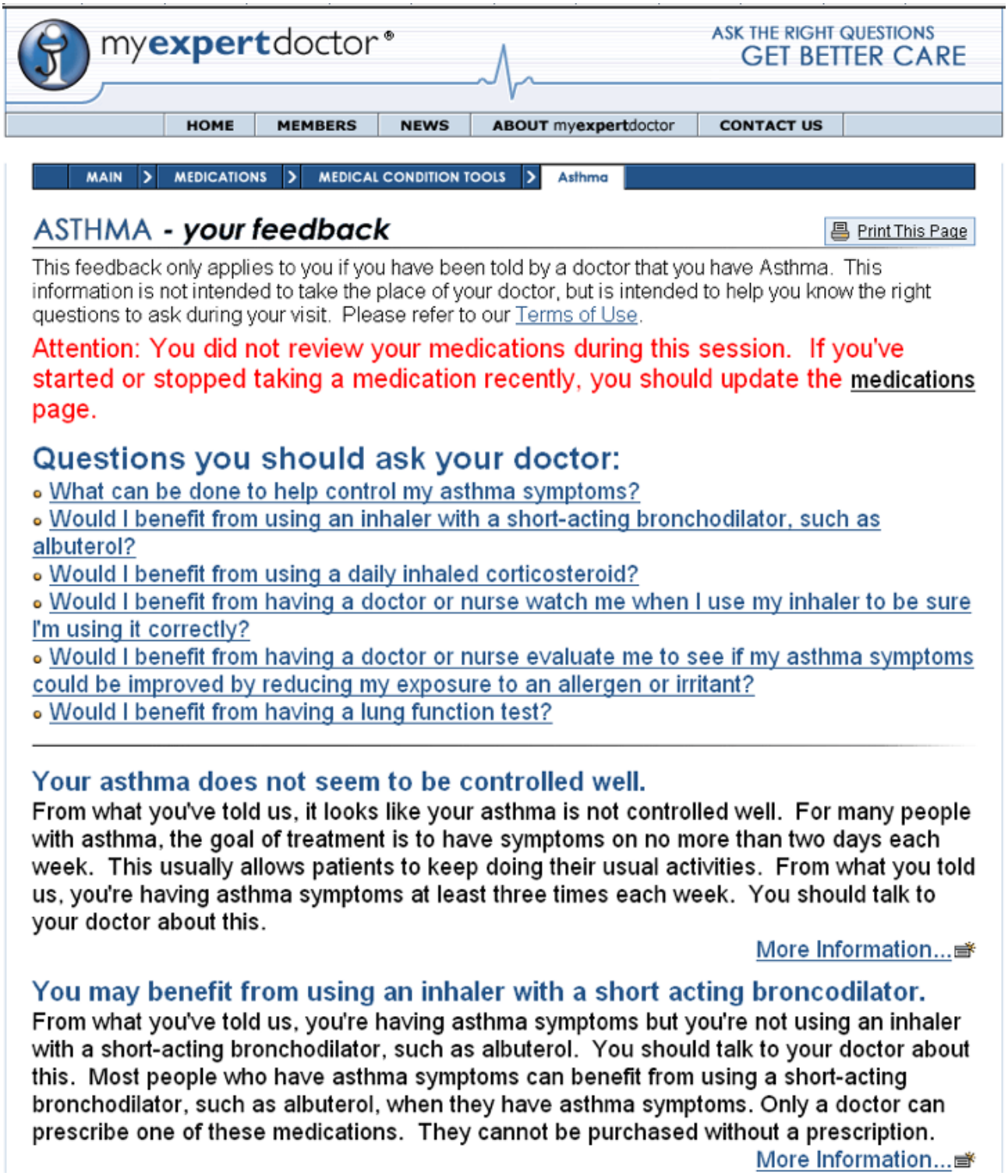
Introduction

Asthma is one of the most common chronic conditions in the United States, yet it is estimated that approximately three fourths of patients with asthma do not have adequate control [1]. New interventions are needed to improve the care of patients with this condition [2-4]. In 1997 and 2002, the National Heart, Lung, and Blood Institute released guidelines for asthma care [5]. Despite the existence of these guidelines, studies show that health practitioners are not following the recommendations and that there is low compliance and inconsistency in asthma management nationwide [2,6,7].Noncompliance with guidelines can lead to overconsumption of health care resources, increased cost, and increased morbidity [8]. Though patient adherence to medications (eg, corticosteroid inhalers) is partly to blame, lack of asthma control also reflects "clinical inertia," or the tendency of providers to make no treatment changes even though a patient has not achieved a treatment target [9-12]. However, research evidence strongly suggests that patients who ask their health care providers for tests and treatments are more likely to receive them [13,14], though the effect of this strategy on chronic disease management has not been well studied [15].

To test the impact of patients asking their health care providers about tests and treatments they could receive, we developed an interactive website (myexpertdoctor.com) to inform patients about asthma and to provide tailored feedback. The website is designed to be used before a physician visit to help patients know what questions to ask during the visit, which in turn may increase the chance that they receive tests and treatments suggested by evidence-based guidelines (see Figure 1 and the Multimedia Appendix for sample screenshots).

XSL•FO
RenderX

Figure 1. Sample questions for physician and tailored feedback



Computer applications have been used to improve asthma control by improving patient education [16,17], disease monitoring [18,19], and by prompting physicians to practise guideline-concordant care [20-22]. However, we are not aware of any interventions designed to prompt patients to ask questions during provider visits in order to improve the quality of their care. We conducted a qualitative study to understand the effects of a Web-based intervention on the physician-patient relationship and on asthma care [23,24]. Although previous studies have shown that the intervention did, by pointing out deficiencies in the quality of their care, cause users to believe they received worse care [5,26], data are needed to understand more fully the potential effects of such an intervention, in particular, how the intervention can impact doctor-patient communication. That was the goal of the current study.

Methods

Intervention Development

The overall design of the Web-based intervention, which included modules related to various medical conditions,

XSL•FO
RenderX

including migraine and osteoarthritis, has been published elsewhere [5,26]. To review, four steps were used in developing the intervention. First, evidence-based decision rules were identified by reviewing clinical guidelines [5]. Second, a self-report survey was created to measure the adherence to each of the guidelines. Next, tailored feedback and suggested questions were created and prioritized for each guideline (see Table 1 and the Multimedia Appendix). Finally, the questions and feedback were programmed into a secure and reliable website, resulting in a three-step process for study patients accessing the site prior to a physician visit. First, patients were prompted to answer 10-20 questions relating to their asthma and its care. Next, patients received immediate personalized feedback and information about their condition, including a list of suggested questions to ask their physician. And finally, patients were encouraged to use this information and the questions during their upcoming physician visit.

Table 1. Website feedback and suggested question examples

Feedback for Care Not in Keeping with the Guideline	Suggested Question to Ask Physician
You may benefit from using a corticosteroid pill for your asthma. From what you told us, your asthma is "worse than usual." You also told us that you're not taking a corticosteroid pill. You may want to ask your doctor about this. Some doctors choose to add a steroid inhaler instead. Corticosteroid pills are very helpful when asthma symptoms are worse than usual. They can help you to feel better much more quickly. These steroids are not bodybuilding steroids. In general, they should only be taken for a few days. (Click here for more information.)	Since my asthma is worse than usual, would I benefit from taking a steroid medication by mouth?
You may benefit from using an inhaler when you feel short of breath. From what you told us, you have some symptoms from your asthma. You also told us that you don't have an inhaler to use when you feel short of breath. You may want to ask your doctor about this. There are several good inhalers for when you're short of breath, such as albuterol. These inhalers work great to make you feel better quickly. (Click here for more information.)	Would I benefit from using an inhaler with a short-acting bronchodilator, such as albuterol?
You may benefit from using a steroid inhaler. From what you told us, your asthma is not well controlled and you're not using a steroid inhaler. You may want to ask your doctor about this. Steroid inhalers help to prevent you from feeling short of breath. Steroid inhalers are used every day to prevent asthma symptoms. A steroid inhaler is a very effective medicine for asthma. They're not like bodybuilding steroids and are quite safe.	Would I benefit from using a daily inhaled corticosteroid?
You may benefit from using a second inhaler to prevent asthma symptoms. From what you told us, your asthma is not well controlled. Also, you're not using a second type of medicine to prevent asthma symptoms. You may want to ask your doctor about this. Medicines that prevent asthma symptoms come in two types. The first are steroids, such as Beclovent. The second type opens the airways, such as Serevent. Both of these medicines should be used each day, whether you have symptoms or not.	Would I benefit from using a long-acting bronchodilator like salmeterol?
You may benefit from using a peak flow meter at home. It's not always easy to know when your asthma is getting out of control. From what you've told us, you've been to the emergency room at least a couple of times over the past year. Because of that, you might benefit from using a peak flow meter every day. You may want to ask your doctor about this. A peak flow meter is a small plastic tube that you blow in to see how your asthma is doing. That way, if you feel okay, but your peak flow is low, you can make a change before you feel worse. It's also important for you to know what to do depending on your peak flow number. This is also something to discuss with your doctor. This is where an "asthma action plan" comes in handy. This is discussed below.	Would I benefit from using a peak flow meter to monitor my asthma at home?
You may benefit from seeing an asthma specialist. From what you've told us, you've been to the emergency room at least a couple of times over the past year. Because of that, it's important for you to see a specialist once in a while. You haven't seen an asthma specialist in the past year. You may want to ask your doctor about this. A specialist can help you figure out if you need different tests or treatments for your asthma. These doctors include "pulmonologists" and "allergists."	Would I benefit from seeing an asthma specialist at this time?
You could do better to prevent asthma attacks. From what you told us, you're using a medicine to prevent asthma attacks. These are called "controller" medications, such as the Azmacort that you are taking. You are not using your controller medicine every day. You may want to talk to your doctor about this. Controller medications help to prevent you from feeling short of breath. They need to be used every day, even if you feel fine.	How frequently should I be using my controller medication(s)?

The website feedback consisted of three elements: (1) a list of suggested questions for the patient to ask his or her physician, (2) a lay explanation of why the patient should ask the physician these questions (one message for those whose care was in keeping with the guideline [eg, moderate persistent asthma whose medication list included a corticosteroid inhaler] and one for those whose care was not), and (3) links to other websites for further reading and explanations of the suggested topics (sites selected and reviewed by panel of experts). By pointing out areas for potential improvements in care ("quality gaps"), we had a concern that some patients would believe that their provider was not giving them needed care. We expended efforts to make the feedback as neutral as possible with regard to this issue. For example, rather than indicating "you need…," the feedback typically suggested "you may benefit from…." The rationale for the feedback was based on the Chronic Care Model, a theoretical model of chronic illness care in which one of the overarching goals is to foster productive interactions between patients, who actively participate in their care, and providers, who can draw on the expertise of guideline-based reminders [27,28].

Study Design

The study was designed to document the experiences of patients who had used the asthma quality improvement website prior to a visit with their asthma care provider. Because qualitative

XSL•FO
RenderX

methods are useful when conducting exploratory research [23,24], and little is known regarding the impact of these types of interventions on quality of care or on doctor-patient communication, this study employed a semi-structured interview methodology to investigate the range of patient experiences with the website, before and during their doctor visit [29]. The interview guide was created by the research team, including two doctoral-level anthropologists and a board-certified internist (CS). Interview questions focused on patients' impressions of the utility of the website, including their experiences using it and specific website navigation issues, as well as the effect of the tailored feedback from the website on doctor-patient communication and perceived quality of care during patient visits. In addition, physicians were recruited to view the website and subsequently participate in a survey regarding their impressions of the site. Ethical approval for the study was obtained from the Institutional Review Board at The Abacus Group, LLC, Cranston, RI, USA.

Sampling and Recruitment

In response to an advertisement on Google and letters mailed to asthma patients of a large health insurance company in Rhode Island, USA, potential subjects were encouraged to contact research staff and were screened over the phone for inclusion and exclusion criteria. The following inclusion criteria were used: (1) age greater than 21 years, (2) self-reported history of asthma, (3) a planned visit with an asthma care provider in the next 2 months, (4) Internet access at home or work, and (5) asthma care from a primary care provider rather than a specialist (eg, pulmonologist). During the screening phone call we asked patients the date of their next asthma provider visit and set a date for a follow-up phone interview after the visit. The final sample comprised 37 patients, who were mailed an informed consent form. The content of the form was explained over the phone, and subjects were encouraged to call the research team if they had any questions. All 37 subjects returned the consent form by mail before the date of their physician visit. Subjects were reimbursed US $100 for participation in the study.

For the physician sample, a nationally representative database of primary care providers was purchased from a marketing organization. A recruitment letter was mailed to a national random sample of 250 physicians; 26 physicians agreed to participate in the study and completed the anonymous survey on the study website. Physicians were reimbursed US $100 for participation in the study.

Data Collection

A research assistant monitored patients' use of the website to be certain they used it before their physician visit. Emails were sent to every participant 7, 4, and 3 days before the date of the visit, reminding them to use the website. Two of the 37 subjects had not used the website within 72 hours of their planned asthma care provider visit and were called by the research assistant to remind them to use the website before their visit. Prior to the asthma care visit, all 37 subjects used the website, answering questions and receiving personalized feedback. The website and the research assistants encouraged patients to print the individualized feedback and take it with them to the physician visit. After visiting their physician, all patients were contacted and participated in a semi-structured telephone interview that lasted approximately 25 minutes. The interviews were conducted by one of two trained research assistants (DB, HL) using an interview guide created by the research team. Open-ended questions, through the use of follow-up questions and probes, allowed for in-depth exploration of topics such as those related to the individualized feedback provided by the website, how this feedback shaped (if at all) the participants' interactions with their physicians, and whether and how the website and feedback were useful in helping them communicate with their physician [29]. Close-ended questions were also asked and addressed (1) use of the website before the visit, (2) use of the website information during the physician visit, and (3) how use of the website and feedback changed the physician visit, if at all.

Physician data were collected via a brief online survey posted on the study website. Physicians were asked to provide written answers to three close-ended and three open-ended questions regarding their impressions of the website, their perceptions of its usefulness to patients, and their suggestions for improvement.

Data Analysis

Patient interviews were audiorecorded, transcribed, and entered into QSR NVivo qualitative software (version 2.0; QSR International; Melbourne, Australia) in order to facilitate data management and analysis. The transcripts were coded by a doctoral-level anthropologist based on the grounded theory technique, in which codes are drawn from the text and coding involves frequent comparative analysis of the data [30,31]. In order to establish the coding scheme, a random 25% of the interviews were coded by an additional doctoral-level anthropologist and discrepancies in coding were resolved via consensus [24,31,32]. Overall, 108 separate codes were identified. As this was a pilot study, a majority of the interview questions, and therefore the codes, related to patients' experiences using the Internet for medical searches and their impressions and use of specific pages of the website. For purposes of the thematic analysis presented here, we were interested in understanding the experiences of patients using the website and how use of the website impacted their physician visit, with particular attention on the doctor-patient interaction and relationship. For that reason, in the thematic portion of the manuscript, we concentrate on presenting the codes and themes that were germane to this issue. Patients' suggestions for improving the website are presented as well. Physician data were collected online and transferred to an Excel database. All descriptive statistics were calculated using SAS (version 9.1).

Results

Quantitative

The large majority of the patients in this study were female (34/37, 91.9%). Approximately one fifth were under the age of 35 (8/37, 21.6%), and only four patients were over the age of 60. The majority were white (33/37, 89.2%), and 45.9% (17/37) had completed college. Asthma symptoms were experienced three or more times a week by slightly more than four fifths of the patients (30/37, 81.1%). In keeping with data that show younger people are more active in terms of using the Internet to seek out health information [33], 67.6% (25/37) of the patients

XSL•FO
RenderX

in this study indicated using the Internet at least once a week for this purpose.

All patients (N = 37) used the website before the visit with their physician; however, only 36 patients answered the questions regarding use of the website. While most of the patients (26/36, 72%) indicated that they brought a printout of the website feedback to the visit, overall only slightly more than half (21/36, 58%) told their physician that they had visited the website. When asked whether use of the website had influenced the outcomes of their visit in any way, 56% (20/36) answered affirmatively. However, with regard to this response, it is important to note that when patient experiences were explored in-depth, almost all patients indicated that use of the website had influenced the visit in some manner. Most participants rated the website positively: excellent, 25% (9/36); very good, 55.6% (20/36); good, 16.7% (6/36); fair, 2.8% (1/36); and poor, 0%.

With regard to the physician (N = 26) responses to the two close-ended questions pertaining to impressions of the website, a large majority of the physicians rated the website positively: excellent, 11.5% (3/26); very good, 50% (13/26); good, 23.1% (6/26); fair, 11.5% (3/26); and poor, 3.8% (1/26). In addition, slightly more than three quarters (20/26, 76.9%) responded positively to a question asking whether they thought that the website and its feedback would be useful in helping patients receive better health care. No demographic information was collected from physicians.

Qualitative

Overall, patients in this study found that having information from the website positively impacted their interactions with their physicians, and, importantly, while some patients reported some dissatisfaction with the website overall, no patient reported a negative encounter with the physician attributable to use of the website. Physicians also reported positive feelings about the website content, while at the same time offering suggestions for improvement.

Physician Impressions of the Website

Physicians were asked three open-ended questions regarding (1) what they thought of the website's feedback section, (2) what suggestions they had for improving the website, and (3) what other comments they had about the website. There were 22 respondents. Physicians' impressions were very favorable regarding the first question and are not summarized here. However, physicians' responses to the other two items did highlight a number of opportunities for improvement.

The most common comment (n = 10) had to do with providing more specific information about medications:

> *Provide info and precautions concerning specific medications.* [Physician 16]
>
> *You may want to avoid using the term "steroid." Perhaps tell the patient that a medication such as prednisone or Medrol might help their exacerbation. In my experience the term "steroid" can be a big turnoff despite trying to educate the patient.* [Physician 25]

Some physicians (n = 5) believed that the website could be strengthened through inclusion of even more specific information about asthma:

> *[Include] more patient education regarding control of triggering factors.* [Physician 41]
>
> *[Include] some basic information about the physiology of asthma that could explain how medications work, so that patients are more motivated to take their medications when needed.* [Physician 9]

Others (n = 3) suggested the addition of more visual aids:

> *A printout or prototype of an asthma management plan for the patient to review prior to physician consultation may be helpful. Also, charts of predicted peak flow values would likely be more educational for patients than written explanations of normal values.* [Physician 24]
>
> *[Include a] visual/schematic of peak flow, spirometry.* [Physician 55]

A number of other comments fell into no specific category and ranged from encouragement to keep the questions and feedback simple, to providing lists of asthma specialists in the area, to allowing for questions from patients and including an appointment schedule.

Patient Suggestions for Improvement of Website

A total of 37 patients suggested improvements for the website. Their feedback pertained to its not providing enough feedback, not providing new information, not giving feedback that was specific to the user, and not having enough scientific information. For example, some participants (n = 13) found that they wanted more detail in both the questions and the feedback, particularly some of the participants whose asthma was rated by the website as being "under control."

> *[The website] didn't come up with any questions or anything to ask my doctor. It pretty much just said, "Oh yeah, your asthma is under control."* [Patient 336: female, age 34]
>
> *I expected more detailed questions [in the feedback section]. It seemed pretty general.* [Patient 309: female, age 36]

Another type of comment had to do with the website's feedback not providing enough information that was new (n = 10). These participants were knowledgeable about their condition because they had either had the condition for many years or had spent time researching it using either the Internet or other means.

> *I've done so much research already that most of what it was telling me, I already knew.* [Patient 338: female, age 33]
>
> *I go to other websites and read about it to learn about it and um...I think if a website is gonna be used to help people ask questions, I think there should be a little bit more information about the asthma itself on the site.* [Patient 321: female, age 35]

In addition, some users (n = 5) perceived the feedback as not being specific enough to their own situation. While this opinion

XSL•FO
RenderX

was only held by a minority of participants, these users felt that the feedback either too closely mimicked the original questions that had been asked or that the feedback was too general to fit their specific situation.

> *It may be that 90% of the people that have asthma fall into the categories that this would help, but I think I might be in the 10% that would have...a different kind of experience of asthma than the other 90%.* [Patient 359: female, age missing]

> *When it gave me the questions and gave me answers that they think I needed to look into, it basically said the same thing again.* [Patient 326: female, age 41]

A final small group of users (n = 2) perceived the feedback as being not scientific enough, by which they meant not having enough information about current asthma research.

> *I think it would have been good if it [the feedback] had said here is what asthma research is going.* [Patient 355: male, age 30]

> *For me more scientific information may have been useful...like if it [the feedback] gave links to publications for epidemiological studies or drug studies.* [Patient 305: female, age 26]

A majority of the users who expressed some reservations about the website nevertheless found some aspect of it to be helpful, either the feedback, the links to other sites, or simply the encouragement to approach their physician with questions.

Patient Themes

Patients' answers to interview questions regarding the individualized feedback and the impact of the website itself centered around two main themes. First, one common result of visiting the website and subsequently visiting a physician was a positive shift in patient attitudes regarding interactions with their physicians. Patients reported that they had more self-confidence, they talked more during the visit, and they had more confidence in the care they were receiving. A second, related theme focused on patients becoming more actively involved in their own asthma care. They gained a better understanding of their treatment options and of their role in managing the condition.

Theme 1: Positive Shift in Patient Attitudes Toward Interactions with Physician

A majority of patients (20/37, 55.6%) answered "yes" when asked a close-ended question about whether the website had influenced their physician visit in some way. In addition, many of those answering "no" or "unsure" to the same question nevertheless revealed in answers to subsequent open-ended questions that their visit had been positively affected. Common to many patients' statements was an increase in self-confidence with regard to communicating with the physician because of the information and questions from the website.

> *I've been going to this doctor for about 17 years, [but this was] the first time that I've actually gotten anywhere with him as far as changing what he was doing for me.... [The website gave me answers to] a lot of the questions that I had in the back of my mind as to why my doctor wasn't pursuing a different avenue. That was kind of cleared up for me, and [it] also gave me the questions to ask him that seemed to push him in the right direction as far as giving me something on a daily basis instead of the inhaler that I was becoming reliant on.* [Patient 308: female, age 49]

> *[My asking questions and having information] threw my doctor for a loop 'cause it took him by surprise. [Laughs.] But it was a good thing. It was a real positive thing. It actually allowed us to create a stronger working relationship with controlling my asthma.* [Patient 344: female, age 49]

Use of the website also increased the amount of time patients talked during the physician visit. Some patients took the feedback printout with them and used it during their physician visit, while others read it before the visit to remind themselves of the content. Because patients were familiar with the suggested questions on the printout and felt secure about what they wanted to ask their physician, they felt more relaxed and interjected more; this, in turn, had a positive impact on the impression of their interaction with the physician, independent of the answers they received to the questions.

> *[You normally] have the questions that you want to ask, but you don't write them down or if you're really sick then you're not even thinking about it. You're just trying to get there. Your stress level's pretty high. [With the printout], I was relaxed a little bit that I was already ready, and I didn't have a lot of work to do. I can just grab the paper and go. It made it a little bit.... [pause] There was a little more discussion with my doctor about my asthma and a couple of triggers. We talked more.* [Patient 313: female, age 35]

> *If I hadn't had that piece of paper [the feedback form] in my hand, I think I would have just gone as a regular office visit and just sat there and waited for her to make a decision, without contributing. [This time] I was able to speak about the fact that I probably should—or she probably should—look at other means of treatment, and that's different than my usual office visit where I don't make any suggestive contribution to what part of treatment is. I just take it all in.... [This time] I was part of the discussion, part of my treatment, and part of the whole total picture of getting my prescriptions changed.* [Patient 340: female, age 42]

Finally, because the website was able to give them or guide them to specific information about their condition, patients repeatedly mentioned feeling more confident in their understanding of issues related to their asthma and its treatment. This led to what they perceived as improvements in the physician-patient relationship and greater confidence in the care they were receiving. In some cases, the physician's authority was validated by the patient's own research, rather than the research being validated by the physician's authority. This was in part due to the fact that the website and the links it provided gave more information than these participants felt they could glean from their short interactions with their physicians.

XSL•FO
RenderX

> *[The feedback form] gave me an opportunity to start off with some questions.... So I had some kind of say in what was going on. Because when I first went in, he just suggested medications for me and everything else, and I had no clue what the medications were, what they did, side effects, or anything like that. So, at least this way it gave me a little advantage because I knew from reading over the information before I went in kind of what the treatment options were.* [Patient 315: female, age 26]

> *[The information] helped a little bit, that I might actually be building this stronger relationship with my doctor by initiating conversation and questions for him.... I was able to talk to him and feel more confident in the treatment that I'm receiving.* [Patient 313: female, age 35]

> *[The website] listed six or so different [asthma] symptoms, and one of them I don't have and I guess most people do, but I could see that it doesn't have to only be this. So it makes me more confident that what my doctor told me is accurate.... [It] validates what the doctor told me.* [Patient 304: female, age 47]

As can be seen from the quotations above, use of the website promoted affirmative changes in patient attitudes toward their interactions with their physician. Not only did visit experiences improve from the patients' perspectives, but some patients also indicated that their questions prompted changes in their treatment. In other words, asking specific questions was reported to overcome clinical inertia and result in positive modifications in patients' care.

Theme 2: Patients Becoming More Actively Involved in Their Asthma Care

Participants varied in how they spoke of their knowledge about asthma and their understanding of their role in managing their asthma. A number of patients were generally more knowledgeable about their condition than others, and while some found the information on the website too limited in scope, others thought the website encouraged them to become more active in seeking additional information, either online, in print sources, or from other individuals. Some patients had traditionally relied in part or, to a lesser extent, exclusively on their physicians for information about their treatment. For many of these patients, the information provided in the feedback and from the links to other sites led to an increased understanding of their condition and of how they could become more involved in their own care.

> *[The website feedback included] questions I never thought to ask. One of mine was about asking "Should I be taking steroid medication by mouth?".... I had to do some research on what they considered steroid medications.... [That helped me] to think more about how to take steroids and myself and more questions to ask the doctor.* [Patient 318: female, age 41]

> *[After looking up information on an in-home peak flow meter:] I didn't know that people had those at home at all. I didn't know that that would be something I could do that would help monitor [my asthma]. That was the major thing. I didn't really have a good idea of how sick I was.* [Patient 309: female, age 36]

For both well-educated and less-informed patients, having access to specific information about their condition frequently resulted in the decision to become more actively engaged in their care. What they learned through having used the website led to changes in how they wanted to approach their care. Patients became more aware of treatment opportunities and the need for changes in the frequency or style of their involvement in managing their condition.

> *I liked that it [the website] told me that maybe I should be taking more precautions and looking more into my [asthma], and maybe I'm not controlling it as well as I should, that maybe more attention should be brought to it each day instead of like once a week, twice a week.* [Patient 326: female, age 41]

> *[Asking questions from the feedback sheet] creates a relationship where you're working together to create a plan, and it's not just the doctor creating the plan.... I have more knowledge now to be able to go to him and have him work on me.* [Patient 344: female, age 49]

Patients' responses illustrate increased involvement in their own care as a result of information and feedback attained either directly or indirectly through use of the website. In addition, as was evidenced in the first theme, having patients pursue new treatment alternatives and become more actively involved in their own care led to perceived positive changes in physician behaviors, including medication and monitoring adjustments.

Discussion

Inconsistencies in the implementation of and noncompliance with asthma treatment guidelines contribute to a reduction in health quality for individuals with the condition [1,34]. This study of a novel Web-based intervention investigated the impact of giving patients individualized information that was designed to prompt physician-patient discussions around issues raised by the evidence-based asthma guidelines [35] and improve the quality of care. To date, to our knowledge, no research has been performed on this type of strategy. Analysis of the data collected showed that knowledge gained from the website and its feedback form positively influenced patients' interactions with their physicians, their knowledge about asthma, and their feelings of responsibility for managing their condition. In addition, patients accepted and enjoyed using the technology to assist with their asthma care, and physicians had positive impressions of the site and its potential to improve care.

Our qualitative results validate the findings of many quantitative studies that have examined similar issues using survey data. Our previous research has shown that patients have a need for the information that is provided on the website. For example, in 2001, we asked 300 primary care patients if they had ever used or had any interest in using the Internet to "find out what questions you should ask your doctors when you see them" [36].

XSL•FO
RenderX

Nearly 60% of the patients were interested in using the Internet for this purpose, and fewer than 30% of the patients had ever found this information on the Internet. It is therefore not surprising that such a high percentage of patients in this study were satisfied with the asthma website. In a similar survey-based study, in which patients were observed searching the Internet for health information before a doctor visit, most (90%) reported feeling more satisfied with their visit than with previous visits because of the Internet use [37]. In that study, patients were trained to use a specific website with links to specifically chosen patient education websites. Those positive findings, coupled with our current findings, suggest that a guided Internet search experience can be quite acceptable to patients and can improve satisfaction with subsequent doctor visits [37].

We hoped that by providing patients with a small number of evidence-based questions, they would have an enhanced patient-provider experience. A previous review of the effects of the Internet on doctor-patient communication identified some concern that physicians may be annoyed by patients bringing in information from the Internet, and that this may possibly harm the doctor-patient relationship [38]. For example, only 15% of 168 physicians surveyed believed that the Internet would improve their relationship with patients, while 49% felt it may harm it [39]. Our results generally point toward the intervention helping patients to communicate better with their physician in order to become more of a partner in their own care. No patient mentioned having a negative interaction with the physician as a result of having used the website. Additionally, the physicians themselves rated the website positively and indicated that they thought the website would be useful in helping to improve patient health care.

We were pleased that subjects viewed the website as reinforcing the care they were receiving from their physician, rather than causing them to question it and potentially undermining their belief in their physician. Our findings are consistent with two other published studies of the website, which observed that the website feedback had no adverse impact on patient perceptions of overall quality of care from a physician [25] or care during a physician visit [27]. This is also consistent with the findings of Kivits [40], who observed that patients viewed Internet health information as complementing, rather than opposing, information from their doctor.

Limitations

Data in this exploratory study were collected from a convenience sample of individuals—the majority of the patient participants were white, female, and over age 35. Because of the sampling method, the results may not be generalizable to patients with asthma, the general patient population, or the physician population. In addition, patients were prompted several times to use the website before their provider visit, including phone calls, which would not occur if the website were implemented in a nonstudy setting. However, although research in the area of using websites to impact physician-patient interactions is new, previous research does support patients' positive reactions to prompts designed to further physician-patient communication [41,42].

Because some of the patients' physicians may have seen the printout from the website and some patients may have told their providers that they were participating in a study, a Hawthorne effect cannot be ruled out. Nevertheless, as the intention of the study was to have patients use the information from the website to prompt discussions with their physicians through the use of the feedback form, we judged the bias attributable to this effect to be limited. Although some providers may have known that their patients were study participants and therefore may have paid more attention to the patients' questions, none of the data collected from the patients themselves indicated that this was the case.

Finally, although patients reported changes in the interactions with their physicians, these may not have translated into changes in patient care. Further research needs to be conducted on whether the website is successful in improving patient outcomes.

Implications for Practice

The Internet has great capacity to positively influence health care, and there is good evidence that it is already an entrenched part of the medical landscape. In the United States, 60% of adults have Internet access, and over 80% of patients with Internet access have searched for health information online [33]. Similarly, we have observed in a survey of 330 primary care patients that most (62.1%) patients felt that their doctor should "recommend specific websites where I can learn more about my health and health care" [43]. However, once physicians encourage patients to use the Internet, patients are likely to hold them accountable for discussions of the information they find. In a separate study, we observed that patients whose physicians did not discuss information gleaned from a tailored-message computer program, much like the website we have designed, were significantly less satisfied with their visit [44]. This suggests that there are likely to be bumps along the road in making suggestions from websites and discussions of Internet searches a part of routine care.

Nevertheless, having patients ask the most pertinent questions relating to their care may help them get the most out of the normally brief office visit. Having patients access this type of information while in a doctor's office has been shown to be too challenging [45]. And while lists of questions to ask a physician about specific conditions are available online (from the American Heart Association, for example), the website in this study took these types of questions one step further, tailoring the feedback based on individuals' answers and explaining why each feedback question was important. In addition, the feedback page placed questions about controlling the condition at the top because results from a prior study indicated that patients were more likely to ask the questions at the top of the feedback page [26]. It is possible that interventions such as the one in this study may increase the efficiency of brief office visits by allowing patients to access pertinent information at home and come to the visit prepared with a list of individualized, prioritized questions and a greater understanding of why asking these questions is important.

While the website has clear implications for practice, it was not designed to be made available to patients in a physician's office. Rather, it should be viewed instead as a prototype for possible

XSL•FO
RenderX

distribution through a number of channels: (1) managed care organizations, to improve the quality of care they provide and satisfy accreditation requirements from the National Committee on Quality Assurance [46]; (2) employers, to decrease work limitations from chronic diseases such as asthma; and (3) advocacy groups, such as the American Lung Association, to improve the quality of care for their constituents. Being a narrowly focused website, it is a relatively inexpensive intervention that requires minimal maintenance, as guidelines are not published that often and major changes to standards of care occur infrequently. Given the steady increase in Internet access [33], we believe that future versions of myexpertdoctor.com and similar offerings could have a significant and positive impact on asthma care and quality of life among patients with asthma and other chronic conditions.

Conclusions

The present study has given us confidence that the current intervention has the potential to improve the way patients communicate with their provider and that the suggested questions can overcome the clinical inertia of providers. Both physician and patient users of the site provided useful feedback on changes that could be implemented in future versions of the website to make it more effective. The main findings—that use of the website and its feedback form positively influenced patients' interactions with their physicians, their knowledge about asthma, and their feelings of responsibility for managing their condition—all point in the expected direction and suggest that the website can improve the quality of care patients receive. We believe that it is essential to give the Internet functionality beyond being a passive, albeit massive, repository of health information. A national study of 4764 adults who used the Internet for health information noted that only one in six believed that the Internet had influenced treatments that they used for a health condition [47]. Although patients may have the potential to learn a great deal, much of the information is beyond their comprehension as it is written at a high reading level and many patients are relatively health illiterate [48,49].

However, empowering patients with specific questions to ask appears to put health information into patients' hands in a way that activates them to be involved in their care. Despite the great number of medically oriented websites, we are not aware of another that provides patients with evidence-based questions to ask their doctor. Most interactive health websites focus instead on providing tailored risk assessment, such as the Heart to Heart Tool [50], Heart Profilers on the American Heart Association website, RealAge, and WebMD. Ongoing studies are evaluating the effect of the website we have designed on patient health outcomes. Given the steady increase in Internet access, we believe that if the current intervention proves to be effective, it may have a significant impact on the control of asthma, as well as other chronic medical conditions.

Acknowledgments

The research was funded via a National Heart Lung and Blood Institute Small Business Innovation Research Grant to Green Lights, Inc.

Conflicts of Interest

Dr. Sciamanna was a principal of Green Lights, Inc at the time of the research. Green Lights, Inc is no longer doing business.

Multimedia Appendix

Screenshots of the website [PPT file (MS PowerPoint), 1.8 MB - jmir_v9i1e3_app1.ppt]

References

1. Carlton BG, Lucas DO, Ellis EF, Conboy-Ellis K, Shoheiber O, Stempel DA. The status of asthma control and asthma prescribing practices in the United States: results of a large prospective asthma control survey of primary care practices. J Asthma 2005 Sep;42(7):529-535. [Medline: 16169784] [doi: 10.1081/JAS-67000]
2. Meng YY, Leung KM, Berkbigler D, Halbert RJ, Legorreta AP. Compliance with US asthma management guidelines and specialty care: a regional variation or national concern? J Eval Clin Pract 1999 May;5(2):213-221. [Medline: 10471231] [doi: 10.1046/j.1365-2753.1999.00177.x]
3. Rundall TG, Shortell SM, Wang MC, Casalino L, Bodenheimer T, Gillies RR, et al. As good as it gets? Chronic care management in nine leading US physician organisations. BMJ 2002 Oct 26;325(7370):958-961 [FREE Full text] [Medline: 12399351] [doi: 10.1136/bmj.325.7370.958]
4. Adams WG, Fuhlbrigge AL, Miller CW, Panek CG, Gi Y, Loane KC, et al. TLC-Asthma: an integrated information system for patient-centered monitoring, case management, and point-of-care decision support. AMIA Annu Symp Proc 2003:1-5. [Medline: 14728122]
5. ; National Asthma Education and Prevention Program. Expert panel report: guidelines for the diagnosis and management of asthma. Update on selected topics 2002. Bethesda, MD: National Heart, Lung, and Blood Institute; 2003. URL: http://www.nhlbi.nih.gov/guidelines/asthma/asthupdt.htm

XSL•FO
RenderX

6. Legorreta AP, Christian-Herman J, O'Connor RD, Hasan MM, Evans R, Leung KM. Compliance with national asthma management guidelines and specialty care: a health maintenance organization experience. Arch Intern Med 1998 Mar 9;158(5):457-464. [Medline: 9508223] [doi: 10.1001/archinte.158.5.457]
7. McGlynn EA, Asch SM, Adams J, Keesey J, Hicks J, Decristofaro A, et al. The quality of health care delivered to adults in the United States. N Engl J Med 2003 Jun 26;348(26):2635-2645 [FREE Full text] [Medline: 12826639] [doi: 10.1056/NEJMsa022615]
8. Stempel DA, Durcannin-Robbins JF, Hedblom EC, Woolf R, Sturm LL, Stempl AB. Drug utilization evaluation identifies costs associated with high use of beta-adrenergic agonists. Ann Allergy Asthma Immunol 1996 Feb;76(2):153-158. [Medline: 8595534]
9. Cabana MD, Rand CS, Powe NR, Wu AW, Wilson MH, Abboud PA, et al. Why don't physicians follow clinical practice guidelines? A framework for improvement. JAMA 1999 Oct 20;282(15):1458-1465 [FREE Full text] [Medline: 10535437] [doi: 10.1001/jama.282.15.1458]
10. Grant RW, Cagliero E, Dubey AK, Gildesgame C, Chueh HC, Barry MJ, et al. Clinical inertia in the management of Type 2 diabetes metabolic risk factors. Diabet Med 2004 Feb;21(2):150-155. [Medline: 14984450] [doi: 10.1111/j.1464-5491.2004.01095.x]
11. O'Connor PJ. Overcome clinical inertia to control systolic blood pressure. Arch Intern Med 2003 Dec;163(22):2677-2678. [Medline: 14662620] [doi: 10.1001/archinte.163.22.2677]
12. Phillips LS, Branch WT, Cook CB, Doyle JP, El-Kebbi IM, Gallina DL, et al. Clinical inertia. Ann Intern Med 2001 Nov 6;135(9):825-834 [FREE Full text] [Medline: 11694107]
13. Kravitz RL, Bell RA, Azari R, Kelly-Reif S, Krupat E, Thom DH. Direct observation of requests for clinical services in office practice: what do patients want and do they get it? Arch Intern Med 2003 Jul 28;163(14):1673-1681. [Medline: 12885682] [doi: 10.1001/archinte.163.14.1673]
14. Kravitz RL, Epstein RM, Feldman MD, Franz CE, Azari R, Wilkes MS, et al. Influence of patients' requests for direct-to-consumer advertised antidepressants: a randomized controlled trial. JAMA 2005 Apr 27;293(16):1995-2002 [FREE Full text] [Medline: 15855433] [doi: 10.1001/jama.293.16.1995]
15. Grol R. Improving the quality of medical care: building bridges among professional pride, payer profit, and patient satisfaction. JAMA 2001 Nov 28;286(20):2578-2585 [FREE Full text] [Medline: 11722272] [doi: 10.1001/jama.286.20.2578]
16. Huss K, Squire EN, Carpenter GB, Smith LJ, Huss RW, Salata K, et al. Effective education of adults with asthma who are allergic to dust mites. J Allergy Clin Immunol 1992 Apr;89(4):836-843. [Medline: 1560167] [doi: 10.1016/0091-6749(92)90439-9]
17. Osman LM, Abdalla MI, Beattie JA, Ross SJ, Russell IT, Friend JA, et al. Reducing hospital admission through computer supported education for asthma patients. Grampian Asthma Study of Integrated Care (GRASSIC). BMJ 1994 Feb 26;308(6928):568-571 [FREE Full text] [Medline: 8093148]
18. Rasmussen LM, Phanareth K, Nolte H, Backer V. Internet-based monitoring of asthma: a long-term, randomized clinical study of 300 asthmatic subjects. J Allergy Clin Immunol 2005 Jun;115(6):1137-1142. [Medline: 15940125] [doi: 10.1016/j.jaci.2005.03.030]
19. Anhøj J, Nielsen L. Quantitative and qualitative usage data of an Internet-based asthma monitoring tool. J Med Internet Res 2004 Sep 3;6(3):e23 [FREE Full text] [Medline: 15471749] [doi: 10.2196/jmir.6.3.e23]
20. Shiffman RN, Freudigman MD, Brandt CA, Liaw Y, Navedo DD. A guideline implementation system using handheld computers for office management of asthma: effects on adherence and patient outcomes. Pediatrics 2000 Apr;105(4 Pt 1):767-773 [FREE Full text] [Medline: 10742318] [doi: 10.1542/peds.105.4.767]
21. McCowan C, Neville RG, Ricketts IW, Warner FC, Hoskins G, Thomas GE. Lessons from a randomized controlled trial designed to evaluate computer decision support software to improve the management of asthma. Med Inform Internet Med 2001;26(3):191-201. [Medline: 11706929] [doi: 10.1080/14639230110067890]
22. Kuilboer MM, Van Wijk MAM, Mosseveld M, Van Der Does E, Ponsioen BP, De Jongste JC, et al. Feasibility of AsthmaCritic, a decision-support system for asthma and COPD which generates patient-specific feedback on routinely recorded data in general practice. Fam Pract 2002 Oct;19(5):442-447 [FREE Full text] [Medline: 12356691] [doi: 10.1093/fampra/19.5.442]
23. Padgett DK. Qualitative methods in social work research: challenges and rewards. Thousand Oaks, CA: Sage; 2004.
24. Miles MB, Huberman AM. Qualitative data analysis. Thousand Oaks: Sage; 1994.
25. Sciamanna CN, Harrold LR, Manocchia M, Walker NJ, Mui S. The effect of web-based, personalized, osteoarthritis quality improvement feedback on patient satisfaction with osteoarthritis care. Am J Med Qual 2005 May;20(3):127-137. [Medline: 15951518] [doi: 10.1177/1062860605274518]
26. Sciamanna CN, Nicholson RA, Lofland JH, Manocchia M, Mui S, Hartman CW. Effects of a Website designed to improve the management of migraines. Headache 2006 Jan;46(1):92-100. [Medline: 16412156] [doi: 10.1111/j.1526-4610.2006.00312.x]
27. Wagner EH, Glasgow RE, Davis C, Bonomi AE, Provost L, McCulloch D, et al. Quality improvement in chronic illness care: a collaborative approach. Jt Comm J Qual Improv 2001 Feb;27(2):63-80. [Medline: 11221012]

XSL•FO
RenderX

28. Wagner EH. Chronic disease management: what will it take to improve care for chronic illness? Eff Clin Pract 1998;1(1):2-4 [FREE Full text] [Medline: 10345255]
29. Rubin HJ, Rubin IS. Qualitative interviewing: the art of hearing data. Thousand Oaks, CA: Sage; 2005.
30. Strauss A, Corbin J. Basics of qualitative research: techniques and procedures for developing grounded theory. 2nd edition. Thousand Oaks, CA: Sage; 1998.
31. Pope C, Ziebland S, Mays N. Qualitative research in health care. Analysing qualitative data. BMJ 2000 Jan 8;320(7227):114-116 [FREE Full text] [Medline: 10625273] [doi: 10.1136/bmj.320.7227.114]
32. Maxwell JA. Qualitative research design: an interactive approach. 2nd edition. Thousand Oaks, CA: Sage; 2005.
33. Fox S. Health information online: eight in ten Internet users have looked for health information online, with increased interest in diet, fitness, drugs, health insurance, experimental treatments, and particular doctors and hospitals. Washington, DC: Pew Internet & American Life Project; 2005. URL: http://www.pewinternet.org/PPF/r/156/report_display.asp
34. Boulet LP, Phillips R, O'byrne P, Becker A. Evaluation of asthma control by physicians and patients: comparison with current guidelines. Can Respir J 2002 Nov;9(6):417-423. [Medline: 12522488]
35. ; National Asthma Education and Prevention Program. Expert panel report: guidelines for the diagnosis and management of asthma. Update on selected topics 2002. J Allergy Clin Immunol 2002 Nov;110(5 Suppl) S141-S219. [Medline: 12542074]
36. Sciamanna CN, Clark MA, Houston TK, Diaz JA. Unmet needs of primary care patients in using the Internet for health-related activities. J Med Internet Res 2002 Dec 31;4(3):E19 [FREE Full text] [Medline: 12554550] [doi: 10.2196/jmir.4.3.e19]
37. Helwig AL, Lovelle A, Guse CE, Gottlieb MS. An office-based Internet patient education system: a pilot study. J Fam Pract 1999 Feb;48(2):123-127. [Medline: 10037543]
38. Blanch DC, Sciamanna CN, Lawless HK, Diaz JA. Effect of the internet on the doctor-patient relationship: a review of the literature. Journal on Information Technology in Healthcare 2005;3(3):179-201.
39. Eberhart-Phillips J, Hall K, Herbison GP, Jenkins S, Lambert J, Ng R, et al. Internet use amongst New Zealand general practitioners. N Z Med J 2000 Apr 28;113(1108):135-137. [Medline: 10872432]
40. Kivits J. Informed patients and the Internet: a mediated context for consultations with health professionals. J Health Psychol 2006 Mar;11(2):269-282 [FREE Full text]
41. Fleissig A, Glasser B, Lloyd M. Encouraging out-patients to make the most of their first hospital appointment: to what extent can a written prompt help patients get the information they want? Patient Educ Couns 1999 Sep;38(1):69-79. [Medline: 14528573] [doi: 10.1016/S0738-3991(98)00124-4]
42. Clayton J, Butow P, Tattersall M, Chye R, Noel M, Davis JM, et al. Asking questions can help: development and preliminary evaluation of a question prompt list for palliative care patients. Br J Cancer 2003 Dec 1;89(11):2069-2077. [Medline: 14647140] [doi: 10.1038/sj.bjc.6601380]
43. Diaz JA, Sciamanna CN, Evangelou E, Stamp MJ, Ferguson T. Brief report: What types of Internet guidance do patients want from their physicians? J Gen Intern Med 2005 Aug;20(8):683-685 [FREE Full text] [Medline: 16050874] [doi: 10.1111/j.1525-1497.2005.0115.x]
44. Sciamanna CN, Novak SP, Houston TK, Gramling R, Marcus BH. Visit satisfaction and tailored health behavior communications in primary care. Am J Prev Med 2004 Jun;26(5):426-430. [Medline: 15165659] [doi: 10.1016/j.amepre.2004.02.003]
45. Sciamanna CN, Marcus BH, Goldstein MG, Lawrence K, Swartz S, Bock B, et al. Feasibility of incorporating computer-tailored health behaviour communications in primary care settings. Inform Prim Care 2004;12(1):40-48. [Medline: 15140352]
46. ; National Committee for Quality Assurance. The State of Health Care Quality: 2005. Washington, DC: National Committee for Quality Assurance; 2005. p. 1-76 URL: http://www.ncqa.org/Docs/SOHCQ_2005.pdf
47. Baker L, Wagner TH, Singer S, Bundorf MK. Use of the Internet and e-mail for health care information: results from a national survey. JAMA 2003 May 14;289(18):2400-2406 [FREE Full text] [Medline: 12746364] [doi: 10.1001/jama.289.18.2400]
48. 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 May;285(20):2612-2621 [FREE Full text] [Medline: 11368735] [doi: 10.1001/jama.285.20.2612]
49. ; Committee on Health Literacy. Health literacy: a prescription to end confusion. Washington, DC: Institute of Medicine; 2004.
50. Pignone M, Sheridan SL, Lee YZ, Kuo J, Phillips C, Mulrow C, et al. Heart to Heart: a computerized decision aid for assessment of coronary heart disease risk and the impact of risk-reduction interventions for primary prevention. Prev Cardiol 2004;7(1):26-33. [Medline: 15010625]

XSL•FO
RenderX

submitted 05.06.06; peer-reviewed by S Krishna; comments to author 08.08.06; revised version received 12.12.06; accepted 12.12.06; published 07.02.07.

Please cite as:
Hartmann CW, Sciamanna CN, Blanch DC, Mui S, Lawless H, Manocchia M, Rosen RK, Pietropaoli A
A Website to Improve Asthma Care by Suggesting Patient Questions for Physicians: Qualitative Analysis of User Experiences
J Med Internet Res 2007;9(1):e3
URL: http://www.jmir.org/2007/1/e3/
doi:10.2196/jmir.9.1.e3
PMID:17478412

© Christine W Hartmann, Christopher N Sciamanna, Danielle C Blanch, Sarah Mui, Heather Lawless, Michael Manocchia, Rochelle K Rosen, Anthony Pietropaoli. Originally published in the Journal of Medical Internet Research (http://www.jmir.org), 07.02.2007. Except where otherwise noted, articles published in the Journal of Medical Internet Research are distributed under the terms of the Creative Commons Attribution License (http://www.creativecommons.org/licenses/by/2.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited, including full bibliographic details and the URL (see "please cite as" above), and this statement is included.

XSL•FO
RenderX

Original Paper

Term Identification Methods for Consumer Health Vocabulary Development

Qing T Zeng[1], PhD; Tony Tse[2], PhD; Guy Divita[2,3], MS; Alla Keselman[2,4], PhD; Jon Crowell[1], MS; Allen C Browne[2], MS; Sergey Goryachev[1], MS; Long Ngo[1], PhD

[1]Decision Systems Group, Brigham and Women's Hospital, Harvard Medical School, Boston, MA, USA

[2]LHNCBC, National Library of Medicine, NIH, DHHS, Bethesda, MD, USA

[3]Management Systems Designers, Inc, Fairfax, VA, USA

[4]Aquilent, Inc, Laurel, MD, USA

Corresponding Author:
Qing T Zeng, PhD
Harvard Medical School
Decision Systems Group
Brigham and Women's Hospital
Thorn 304, 75 Francis Street
Boston, MA 02115
USA
Phone: +1 617 732 7694
Fax: +1 617 739 3672
Email: qzeng@dsg.harvard.edu

Abstract

Background: The development of consumer health information applications such as health education websites has motivated the research on consumer health vocabulary (CHV). Term identification is a critical task in vocabulary development. Because of the heterogeneity and ambiguity of consumer expressions, term identification for CHV is more challenging than for professional health vocabularies.

Objective: For the development of a CHV, we explored several term identification methods, including collaborative human review and automated term recognition methods.

Methods: A set of criteria was established to ensure consistency in the collaborative review, which analyzed 1893 strings. Using the results from the human review, we tested two automated methods—C-value formula and a logistic regression model.

Results: The study identified 753 consumer terms and found the logistic regression model to be highly effective for CHV term identification (area under the receiver operating characteristic curve = 95.5%).

Conclusions: The collaborative human review and logistic regression methods were effective for identifying terms for CHV development.

(J Med Internet Res 2007;9(1):e4) doi:10.2196/jmir.9.1.e4

KEYWORDS

Consumer health information; vocabulary; natural language processing

Introduction

Two important steps in vocabulary development are (1) the identification of candidate strings (ie, words or phrases) in a domain and (2) the determination of which of these should be included in a vocabulary as "valid" terms, also called "termhood determination." Health vocabulary development, which has a long history, requires significant effort for collecting candidate terms and determining termhood [1]. While vocabularies such as SNOMED (Systematized Nomenclature of Medicine) and ICD-9 (International Classification of Diseases, Ninth Revision) include many health terms, there is no consensus on termhood criteria (ie, what constitutes a "term") [2]. The decision to include terms in a vocabulary is made for a particular domain for certain tasks (eg, indexing or billing). Thus, the review criteria and procedures used by vocabulary developers, which are often not published, inevitably differ. Terms included in health vocabularies also vary significantly. For instance, in the Unified Medical Language System (UMLS), the same concept

XSL•FO
RenderX

is often represented in various source vocabularies by different terms. The terms “head ache” and “cranial pain” are both synonyms of the UMLS concept “headache.” The source vocabulary for “head ache” is DXplain, and the source vocabulary for “cranial pain” is MeSH (medical subject heading).

Research and development of controlled consumer health vocabularies (CHVs) is a relatively new endeavor in the health vocabulary field [3]. In the general biomedical literature, research on consumer understanding of medical words and concepts has focused primarily on relatively short lists of discrete terms in various specialties. In the informatics domain, a few companies (eg, Apelon and WellMed) offer proprietary CHV products, though these products have not been publicly evaluated.

The general goal of our CHV research is to help overcome the vocabulary gap between consumers and health information provided by informatics applications. The specific aim of this paper is to elucidate term identification methods for CHVs. CHV research has largely been driven by the proliferation of health-related materials on the Web, the emergence of electronic personal health records, as well as the growing availability of various consumer health applications (eg, decision support tools). Over the past five years, researchers have found that consumer terms are not well covered by the existing health vocabularies, which mostly represent the language of health professionals [4-9]. Indeed, expressions used by consumers to describe health-related concepts and relationships among such concepts frequently differ on multiple levels (ie, syntactic, conceptual, and explanatory) from those of professionals. Thus, consumer health informatics research and application development will benefit from the development of CHVs.

Developing and validating a comprehensive CHV is challenging because “consumers” constitute a plethora of highly diverse groups. Further, individuals uniquely acquire health-related terms and concepts from formal and informal sources (eg, media exposure) and from personal experiences. Nevertheless, there is strong evidence of the stability of lay health language among particular populations, for specific tasks [3].

We have been working on an open access and collaborative (OAC) CHV project. The first step in creating the OAC CHV was to identify consumer terms since surface forms, represented as strings in written text, are more tractable than concepts (ie, underlying meanings) or semantic relations, both of which require in-depth understanding of term usage, rhetorical intent, and explanatory models. Because consumer terms are heterogeneous and even less well defined than professional terms [10], the termhood determination task proved to be particularly challenging. Our term identification effort has been guided by two principles:

1. CHVs consist of actual terms commonly used by consumers (in any particular discourse group).

2. CHV terms must allow for computer processing of consumer language.

Since many professional health vocabulary terms are already used by consumers, though in some cases with different or broader semantics (eg, “diabetes” for diabetes mellitus, types 1 and 2), we focused on consumer terms not yet represented in existing vocabularies (eg, “broken finger” for any type of fracture in the “distal,” “middle,” or “proximal phalanges”).

Because the number of candidate strings is often very large in any domain, researchers have explored the use of corpus-based automated term recognition (ATR) methods for extracting the most promising strings for human review from domain-specific documents [11, 12]. ATRs vary from statistical or information theory–based approaches (eg, *t* test) [13] to syntax-based methods (eg, noun phrase extraction and context analysis) [14] and hybrid mechanisms (eg, C-value formula) [15, 16]. Both the *t* test and the C-value formula have been used successfully in termhood determination. Such studies reinforce the general notion that strings typically considered as terms share some common characteristics, such as words in a term tend to occur more frequently together, terms are often noun phrases, and terms may be part of several longer strings.

In the biomedical domain, ATR methods have been applied to Medline literature [17] and clinical reports [15]. While most ATR methods outside the biomedical domain were designed to be general purpose, biomedical ATR methods tend to be more narrowly focused [18]. The type of terms targeted by ATR vary, including gene and protein names in a number of recent studies [18-21].

In this study, we first identified CHV terms through collaborative review of strings derived from query logs of a consumer health site [22]. Because of the considerable variability in lay health expressions, standardized review criteria and procedures to ensure consistency in selecting CHV terms were developed. After obtaining the human-reviewed n-grams (ie, n word strings), we experimented with two ATR methods: logistic regression and the C-value formula. The initial features used in the regression model were informed by existing ATR methods, in particular, the C-value model [16] and the termhood formula proposed by Wermter and Hahn [12]. We also evaluated the popular C-value method.

Our use of ATRs in this study differs from that in prior studies in the biomedical domain in two aspects: (1) short phrases from query logs were used as the text corpus rather than entire sentences from full-text sources, and (2) “new” CHV terms, not yet part of existing vocabularies, were identified rather than “pre-existing” terms such as UMLS terms.

Methods

The term identification study had three components:

1. Candidate string extraction from a query log data set of terms that could not be mapped to UMLS
2. Collaborative manual review of a subset of the candidate strings and identification of CHV terms
3. Application of ATR methods (the C-value formula and logistic regression models) to human-reviewed CHV terms

Candidate String Extraction

We obtained a set of query log files [22] from the MedlinePlus site covering the period from October 2002 to October 2003,

XSL•FO
RenderX

courtesy of the National Library of Medicine (NLM). The log data were preprocessed to filter out all queries that were not in English, appeared to be machine generated (eg, very large numbers of queries from the same IP address within a minute), and that were redundant (ie, from the same host at time intervals of less than 5 minutes).

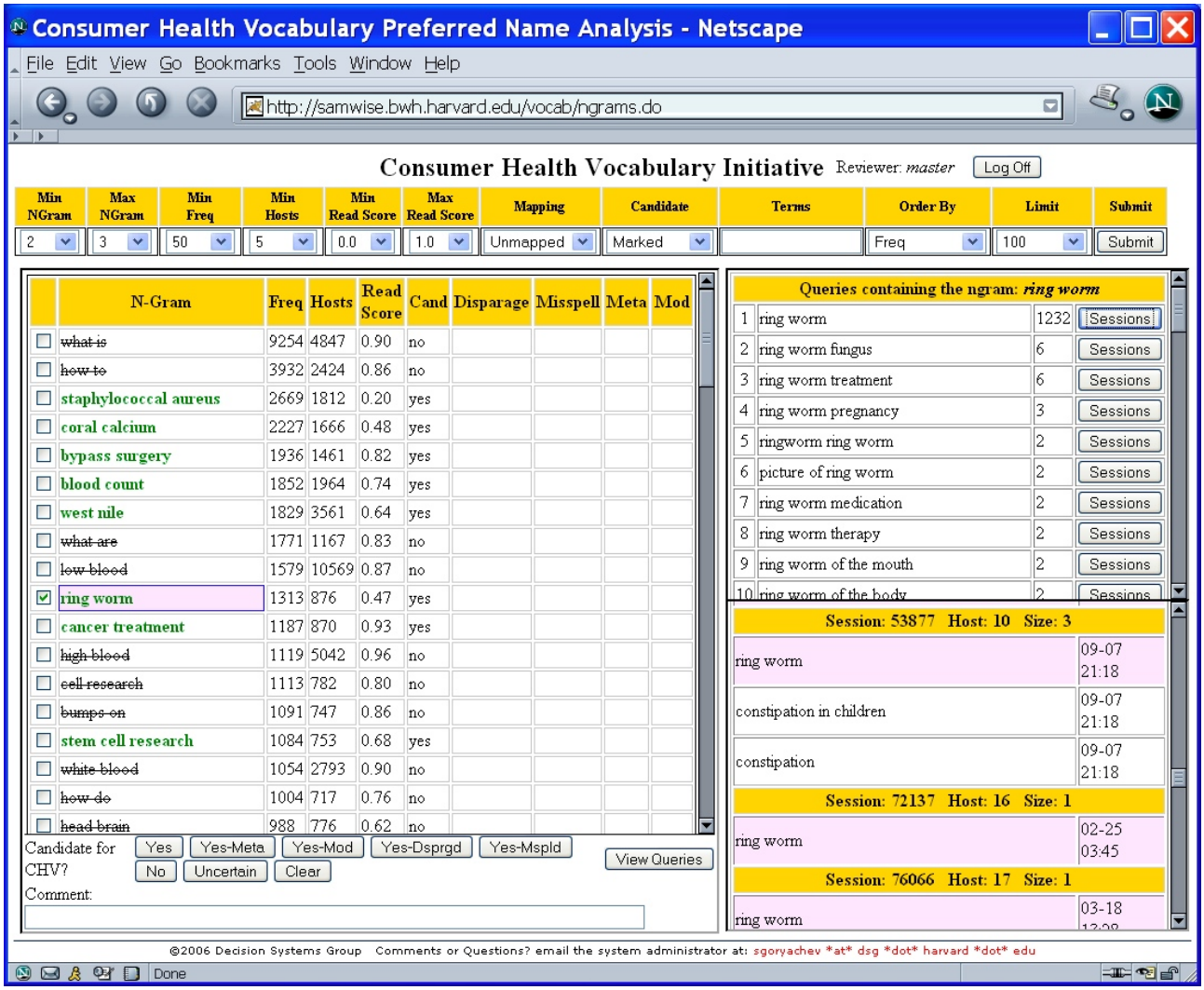
The preprocessed queries were then mapped to the 2004AA version of the UMLS Metathesaurus using lexical methods (ie, removing non-alphanumeric symbols, stemming, normalization, and truncation). Queries that did not map to the UMLS Metathesaurus were broken into n-grams. N-grams that matched terms in the Metathesaurus were removed, and the remaining n-grams were collected into sets by frequency and number of words.

We used n-gram analysis to find candidate terms from unmapped query strings. The n-gram analysis uses the frequencies of n-grams and text fragments of n words in a text sample to estimate the likelihood that a string is a potential term. In general, the more frequently an n-gram appears in text documents, the increased likelihood that the n-gram is a "useful" term.

Collaborative Manual Review

Six researchers (first six of the authors) reviewed candidate strings (n-grams) collaboratively. First, each reviewer independently reviewed a subset of the n-grams (n = 1 to 4 and frequency > 50) and voted on whether they should be considered CHV terms. Unanimous votes for n-grams that were reviewed by at least three people were entered as "master" votes. Otherwise, termhood was discussed by the entire group until consensus was reached and a master vote was cast. To support reviewers from geographically distributed locations and to calculate votes, a specially designed Web-based application [23] was utilized (Figure 1).

Figure 1. Application to support collaborative manual review of candidate strings



Through several iterations of votes and discussion, we established the following review criteria:

1. CHV terms should be syntactic constituents or phrases such as a noun phrase or adjective phrase (eg, "bypass surgery" is a phrase, but "fever in" is not). Special attention should be given to noun phrases.
2. CHV terms should have independent semantics and should not only occur as a part of longer valid terms or as a part of wild card searches (eg, [chicken-, small-] "pox vaccine" is not considered a CHV term).
3. CHV terms should be specific to the medical domain (eg, "Google" and "Yahoo" are general words, not CHV terms).

XSL•FO
RenderX

4. CHV terms should function as semantic components in addition to functioning as syntactic components (eg, stop words “the” and “a” as well as empty verbs “make” and “take” are not considered CHV terms).
5. N-grams representing existing UMLS medical concepts are considered to be CHV terms, but CHV terms may represent non-UMLS concepts.
6. Eponymous forms of CHV terms are considered to be CHV terms (eg, “Parkinson’s”).
7. CHV terms may include spelling errors, (eg, “Chron's disease”). These misspelled terms are given the label “disparaged.”
8. Terms with distinct clinical semantics (eg, “result”) are considered to be CHV terms, regardless of ambiguity and/or vagueness in other domains.

We singled out several types of terms for future investigation and assigned special labels to them:

- meta: A term that is usually used to indicate the category/type of information sought or presented (eg, “picture,” “guideline,” and “tutorial”).
- modifier: A term not typically used by itself, but for limiting or qualifying other terms (eg, “sexually” as in “sexually active”).
- relation: A term not typically used by itself, but used to describe relations among concepts (eg, “caused by” and “results in”). We also include the unary relation “not” in this set.

Currently, we consider terms classified as meta and modifier to be CHV terms, but relations are not considered CHV terms.

Once these review criteria were established, researchers double-checked the previously cast master votes for compliance. A second round of discussion resulted in some adjustments to the votes.

Application of Automated Term Recognition (ATR)

We explored the use of two ATR methods to facilitate candidate selection for human review: (1) the C-value method (C loosely stands for “candidate collection”) and (2) logistic regression.

We applied the C-value method to the strings that had already been reviewed. First, the strings were parsed to filter out single-word strings and strings that were not noun phrases. The C-value was calculated using the formula [16] given in Textbox 1.

Textbox 1. The C-value was calculated using this formula

$C\text{-}value(\boldsymbol{a}) = \log_2|\boldsymbol{a}| * f(\boldsymbol{a})$ if $\boldsymbol{a}$ is not nested

(When $\boldsymbol{a}$ is a substring of $\boldsymbol{b}$, we refer to $\boldsymbol{a}$ as nested and $\boldsymbol{b}$ as $\boldsymbol{a}$’s nesting string.)

$C\text{-}value(\boldsymbol{a}) = \log_2|\boldsymbol{a}| * (f(\boldsymbol{a}) - 1/p(T\boldsymbol{a}) * \text{sum}(f(\boldsymbol{b})))$ if $\boldsymbol{a}$ is nested

$\boldsymbol{a}$ = candidate string (eg, “failure”)

$\boldsymbol{b}$ = nesting strings (eg, “heart failure”)

$|\boldsymbol{a}|$ = length (number of words) of $\boldsymbol{a}$

$f(\boldsymbol{a})$ = frequency of $\boldsymbol{a}$ in the corpus

$T\boldsymbol{a}$ = set of $\boldsymbol{b}$ that contain $\boldsymbol{a}$

$P(T\boldsymbol{a})$ = number of $\boldsymbol{b}$ in $T\boldsymbol{a}$

$f(\boldsymbol{b})$ = frequency of $\boldsymbol{b}$ in the corpus

To create the logistic regression model that predicts the termhood of a candidate string ***a,*** we explored syntactic category, frequency of occurrence, string length, word count and number, frequency and termhood status of ***a***’s nesting, and nested strings as variables and used the master vote as outcome. Human-reviewed strings were used as the training and testing data sets. The initial feature variables were as follows:

1. part-of-speech (POS) tag (eg, noun or adjective) of the first word
2. POS tag of the last word
3. noun phrase status (ie, yes/no)
4. word count (ie, number of words in ***a)***
5. number of distinct ***a***’s nesting string ***b***
6. number of repeated ***b***
7. percentage of distinct ***b*** that are known valid (UMLS) terms
8. percentage of repeated ***b*** that are known valid (UMLS) terms
9. number of distinct ***a***’s nested string ***c***
10. number of repeated ***c***
11. percentage of distinct ***c*** that are known valid (UMLS) terms
12. percentage of repeated ***c*** that are known valid (UMLS) terms
13. frequency of ***a***
14. number of distinct host ***h*** that ***a*** originated from
15. average number of distinct queries containing ***a*** per host

The frequency distribution of the POS tags (variables 1 and 2) required them to be collapsed into fewer categories for modeling. The original tags came from a Brill-style, rule-based POS tagger developed by Mark Hepple [24]. We first transformed them into a smaller set of tags used by the UMLS SPECIALIST Lexicon of the National Library of Medicine (NLM) [25]. (Details of the transformation rules can be found in [26].) Several tags appeared with low frequency and were then merged: the tags AUXILARY and MODAL were merged with VERB, and the tags CONJUNCTION, DETERMINER,

XSL•FO
RenderX

NUMBER, SYM, UNKNOWN, PRONOUN, and PREP were merged into a new category, OTHER.

The continuous variables (variables 4 to 15) were dichotomized based on the median value. The dichotomized variables were used in the logistic regression to predict or explain the probability of having a term voted "yes" for termhood.

The logistic regression model building was carried out by a stepwise procedure. After calculating the odds ratio estimates, most of the variables were dropped. The remaining variables 1, 2, 3, 6, 10, and 15 were represented in the regression formula as FirstPOS, LastPOS, np_value, repeat_sup_gt_median, repeat_sub_gt_median, and distinct_perhost_gt_median.

For both the C-value formula and the regression model, we calculated the sensitivity and specificity at different thresholds to create the receiver operating characteristic (ROC) curves. To estimate the area under the ROC curve for the logistic regression, we used the c-statistic [27] (note that this is not the same as C-value). It has the following meaning. From the final multivariable logistic regression model, the predicted probability of the termhood voted "yes" can be computed for each term. For any two terms, one with vote "yes" and one with vote "no," if the predicted probability for vote "yes" is higher than the predicted probability for vote "no," then we have a concordant pair. If the predicted probability of vote "no" is higher, then we have a discordant pair. If the pair is neither concordant nor discordant, then it is tied. Let T be the total number of all possible yes-no pairs of all terms. Let C be the number of concordant pairs, and D the number of discordant pairs. The c-statistic is calculated as $c = (C + 0.5(T - C - D)) / T$.

Results

We identified 18454 candidate n-grams (n = 1 to 5); 7967 were reviewed by at least one reviewer, and 1893 distinct n-grams received master votes (Table 1). Among the n-grams with master votes, 23 were meta, 39 were modifier, and 5 were relation.

Table 1. Number of n-grams with master votes and number of n-grams voted as CHV terms

N-gram	Number of Master Votes	Number of CHV Terms
1-gram	379	261
2-gram	1101	303
3-gram	356	154
4-gram	57	35
Total	**1893**	**753**

Figure 2. The logistic regression model

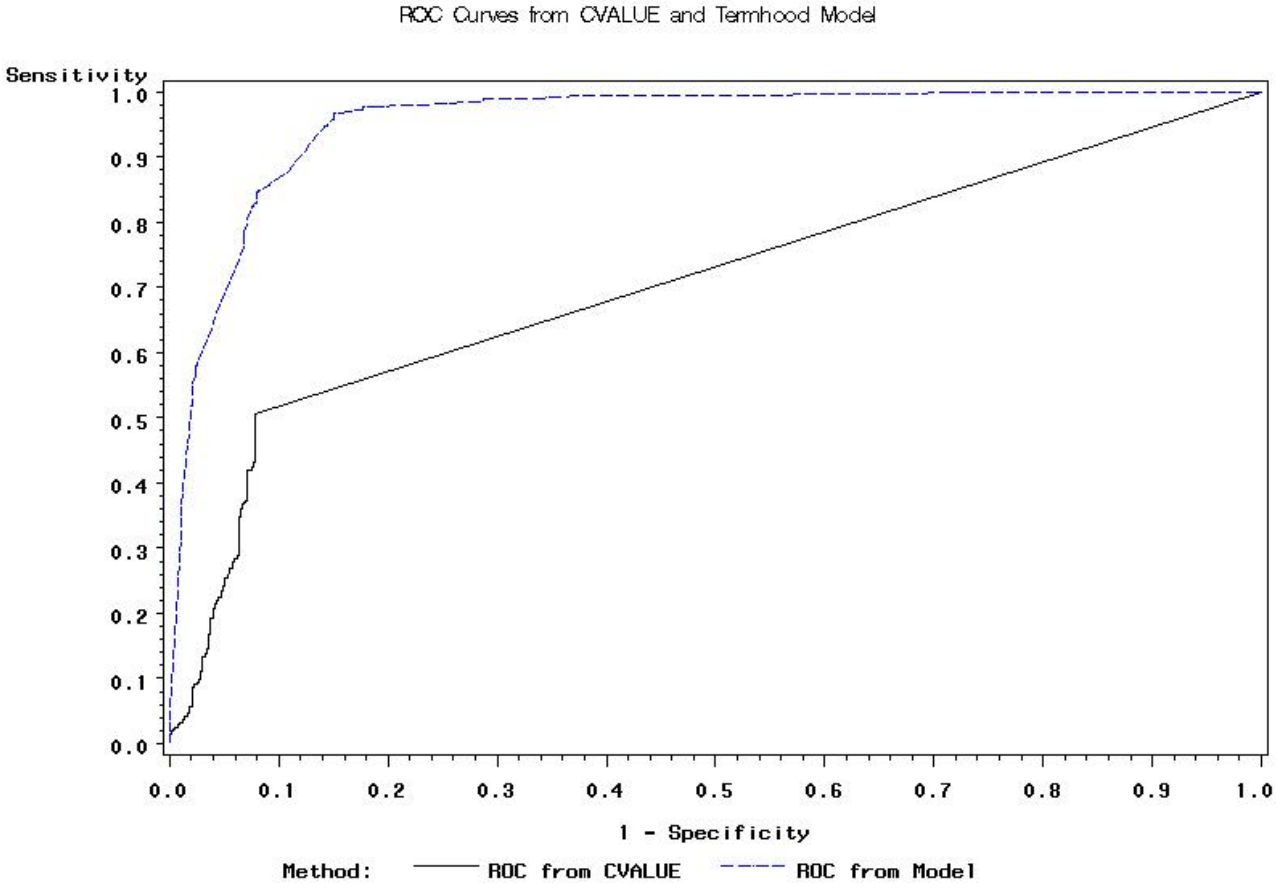
$$\log\left(\frac{P(vote=yes)}{P(vote=no)}\right) = -0.7836 + 0.7541 * FirstPOS_ADJECTIVE - 1.3722 * FirstPOS_ADVERB$$
$$+ 0.3541 * FirstPOS_NOUN + 1.4182 * FirstPOS_VERB - 0.7722 * LastPOS_ADJECTIVE$$
$$+ 2.2576 * LastPOS_ADVERB + 0.0285 * LastNOUN + 0.6038 * LastPOS_VERB + 1.2899 * NP_VALUE$$
$$+ 1.0475 * REPEAT_SUP_GREATER_MEDIAN + 0.8417 * REPEAT_SUB_GREATER_MEDIAN$$
$$0.8422 * DISTINCT_PERHOST_GREATER_THAN_MEDIAN$$

The logistic regression model is shown in Figure 2. In this logistic regression model, syntactic information (first 9 variables) and nesting pattern (last 3 variables) determine the termhood. The importance of syntactic information has long been recognized by models like the C-value. Conspicuously, word count and frequency are missing from our model, though longer and more frequent strings are more likely to be considered terms. To a large extent, length and frequency are reflected by the nesting patterns: very short strings are likely to be part of many nesting strings, and less frequent strings are likely to be coincidental combinations of more common words, meaning that it would have more nested strings.

The ROC curves for C-value and the regression model are shown in Figure 3. The area under the ROC curve (AUC) is 70.9% for the C-value method and 95.5% for the regression model. Higher AUC signifies increased distinguishing power: 100% = perfect discriminative ability, 50% = no ability, < 50% = predications were made in the wrong direction. Thus, the AUC results suggest the regression model to be very effective and better than the C-value for identifying CHV terms.

XSL•FO
RenderX

Figure 3. Curves for C-value and the regression model



Discussion

This paper reports on several term identification methods for the OAC CHV project. We established a set of criteria and procedures to conduct a manual review, resulting in multiple reviewers reaching consensus on 1893 n-grams, including identification of 753 new terms for inclusion in the OAC CHV that were not in the 2004AC version of UMLS.

The OAC termhood criteria were established collaboratively, reflecting the reviewers' backgrounds in several different fields: controlled vocabulary, health informatics, linguistics, cognitive science, and computer science. While the OAC termhood criteria could be further refined and termhood criteria for health vocabularies are often not published, we believe publishing such criteria could benefit vocabulary research. For instance, many articles evaluate vocabularies and study methods of mapping one vocabulary to another [28-31]. These evaluations and mapping methods could be better guided by the termhood criteria of target vocabularies.

In CHV research, the termhood issue is of particular importance because there has been limited discussion and little consensus on what should be considered a consumer term. Is "sun poisoning" an acceptable term? How about "skin conditions?" As was pointed out in the Introduction, health professional vocabularies do not always agree on the termhood of a phrase. Consumer expressions, however, require more scrutiny because it is harder to determine their semantics and contexts of usage.

We tested two ATR methods (C-value and logistic regression) on the human-reviewed n-grams. The C-value was useful for determining termhood, though it did not have high distinguishing power (AUC = 70.9%). The AUC for the logistic regression model was 95.5%, which is fairly satisfactory.

These results suggest that a specially fitted logistic regression model is better suited than the generic C-value method for the task of identifying CHV terms according to our criteria. The C-value method's performance problem was partially caused by issues unique to this data set, among them the inclusion of infrequent misspellings and the high frequency of most candidates, which made frequency a less reliable predicator. The imperfection in noun-phrase parsing is not unexpected, though the relatively short query string posed a greater challenge for parsing. Like many vocabularies, OAC includes strings that are single words and are not noun phrases, while C-value is typically calculated for multiword noun phrases.

The logistic regression model demonstrated excellent suitability for OAC termhood determination. It may have to be altered to be used with other corpora or for other types of vocabularies due to the particularities of query-based corpus attributes such as the short length of the documents. Nonetheless, training of predictive models for a particular corpus and vocabulary is a generalizable strategy. Although general principles exist, the determination of which strings are to be considered legitimate vocabulary terms often depends on the domain and the vocabulary developers' criteria (eg, including verb phrases [15] or not).

XSL•FO
RenderX

The regression model utilizes syntactic and nesting pattern features; both types of features are well-recognized termhood indicators. A concern often raised about CHV research is that the syntax and semantic of consumer phrases are too unruly to be represented in a computable vocabulary. The fact that many consumer phrases have common term characteristics suggests that they are tractable terms.

Our study has several limitations. Because consumer utterances are not readily available as corpora of medical literature or clinical records, we used query logs that contained relatively few complete sentences. Subsequently, this resulted in many POS and noun phrase analysis errors. As well, we only had researchers and not lay consumers review the candidate terms, due to budget and logistic constraints. However, the analysis was based on utterances from queries submitted by tens of thousands of consumers.

Based on the result of this study, we plan to apply the logistic regression model to the candidate n-grams and select those predicted to be terms for human review. We also plan to add the identified CHV terms to OAC. The authors associated with NLM are interested in investigating similar techniques to aid in identifying candidate terms for inclusion into the SPECIALIST Lexicon of the NLM, and for quality control.

Acknowledgments

We thank the National Library of Medicine (NLM) for sharing the MedlinePlus query log data. This work is supported by the National Institutes of Health (NIH) grant R01 LM07222 and by the Intramural Research Program of the NIH, NLM.

Conflicts of Interest

None declared.

References

1. Chute CG. Clinical classification and terminology: some history and current observations. J Am Med Inform Assoc 2000 May;7(3):298-303 [FREE Full text] [Medline: 10833167]
2. Cimino JJ. Desiderata for controlled medical vocabularies in the twenty-first century. Methods Inf Med 1998 Nov;37(4-5):394-403. [Medline: 9865037]
3. Zeng QT, Tse T. Exploring and developing consumer health vocabularies. J Am Med Inform Assoc 2006 Jan;13(1):24-29 [FREE Full text] [Medline: 16221948] [doi: 10.1197/jamia.M1761]
4. Brennan PF, Aronson AR. Towards linking patients and clinical information: detecting UMLS concepts in e-mail. J Biomed Inform 2003 Aug;36(4-5):334-341. [Medline: 14643729] [doi: 10.1016/j.jbi.2003.09.017]
5. Zeng Q, Kogan S, Ash N, Greenes RA, Boxwala AA. Characteristics of consumer terminology for health information retrieval. Methods Inf Med 2002;41(4):289-298. [Medline: 12425240]
6. Mccray AT, Loane RF, Browne AC, Bangalore AK. Terminology issues in user access to Web-based medical information. Proc AMIA Symp 1999:107-111 [FREE Full text] [Medline: 10566330]
7. Tse T. Identifying and Characterizing a “Consumer Medical Vocabulary. ” College Park: University of Maryland; 2003.
8. Patrick TB, Monga HK, Sievert ME, Houston Hall J, Longo DR. Evaluation of controlled vocabulary resources for development of a consumer entry vocabulary for diabetes. J Med Internet Res 2001 Aug 28;3(3):E24 [FREE Full text] [Medline: 11720966] [doi: 10.2196/jmir.3.3.e24]
9. Smith CA, Stavri PZ, Chapman WW. In their own words? A terminological analysis of e-mail to a cancer information service. Proc AMIA Symp 2002:697-701. [Medline: 12463914]
10. Zielstorff RD. Controlled vocabularies for consumer health. J Biomed Inform 2003 Aug;36(4-5):326-333. [Medline: 14643728] [doi: 10.1016/j.jbi.2003.09.015]
11. Krauthammer M, Nenadic G. Term identification in the biomedical literature. J Biomed Inform 2004 Dec;37(6):512-526. [Medline: 15542023] [doi: 10.1016/j.jbi.2004.08.004]
12. Wermter J, Hahn U. Effective grading of termhood in biomedical literature. AMIA Annu Symp Proc 2005:809-813. [Medline: 16779152]
13. Church K, Gale W, Hanks P, Hindle D. Using statistics in lexical analysis. In: Zernik U, editor. Lexical Acquisition: Exploiting On-Line Resources to Build a Lexicon. Hillsdale, NJ: Lawrence Erlbaum Associates, Inc; 1991.
14. Basili R, Pazienza MT, Zanzotto FM. Modelling the syntactic contextual information for term extraction.. In: In: Conference on Recent Advances in Natural Language Processing (RANLP2001). Tzigiv Chark, Bulgaria; Sep 5, 2001.
15. Harris MR, Savova GK, Johnson TM, Chute CG. A term extraction tool for expanding content in the domain of functioning, disability, and health: proof of concept. J Biomed Inform 2003 Aug;36(4-5):250-259. [Medline: 14643720] [doi: 10.1016/j.jbi.2003.09.005]
16. Frantzi KT, Ananiadou S, Mima H. Automatic recognition of multi-word terms: the C-value/NC-value method. Int J on Digital Libraries 2000;3(2):115-130. [doi: 10.1007/s007999900023]
17. Rindflesch TC, Hunter L, Aronson AR. Mining molecular binding terminology from biomedical text. Proc AMIA Symp 1999:127-131 [FREE Full text] [Medline: 10566334]

XSL•FO
RenderX

18. Fundel K, Güttler D, Zimmer R, Apostolakis J. A simple approach for protein name identification: prospects and limits. BMC Bioinformatics 2005;6 Suppl 1(Suppl 1):S15 [FREE Full text] [Medline: 15960827] [doi: 10.1186/1471-2105-6-S1-S15]
19. Saric J, Jensen LJ, Ouzounova R, Rojas I, Bork P. Extraction of regulatory gene/protein networks from Medline. Bioinformatics 2006 Mar 15;22(6):645-650. [Medline: 16046493] [doi: 10.1093/bioinformatics/bti597]
20. Saric J, Jensen LJ, Rojas I. Large-scale extraction of gene regulation for model organisms in an ontological context. In Silico Biol 2005;5(1):21-32 [FREE Full text] [Medline: 15972005]
21. Cohen AM, Hersh WR, Dubay C, Spackman K. Using co-occurrence network structure to extract synonymous gene and protein names from MEDLINE abstracts. BMC Bioinformatics 2005 Apr;6(1):103 [FREE Full text] [Medline: 15847682] [doi: 10.1186/1471-2105-6-103]
22. Zeng QT, Tse T, Crowell J, Divita G, Roth L, Browne AC. Identifying consumer-friendly display (CFD) names for health concepts. AMIA Annu Symp Proc 2005:859-863.
23. Crowell J, Zeng Q, Tse T. A web application to support consumer health vocabulary development. AMIA Annu Symp Proc 2005:932. [Medline: 16779219]
24. Hepple M. Independence and Commitment: Assumptions for Rapid Training and Execution of Rule-based Part-of-Speech Taggers. Proceedings of the 38th Annual Meeting of the Association for Computational Linguistics (ACL-2000). Hong Kong, October 2000. Midtown, NJ: Association for Computational Linguistics; 2000:278-285.
25. Browne A, McCray A, Srinivasan S. The Specialist Lexicon. Report No NLM-LHC-93-1. Bethesda, MD: Lister Hill National Center for Biomedical Communications, National Library of Medicine; 2000.
26. Notes on Tagger Integration. URL: http://mmtx.nlm.nih.gov/taggerNotes.shtml [accessed 2006 Oct 23]
27. Hanley JA, Mcneil BJ. The meaning and use of the area under a receiver operating characteristic (ROC) curve. Radiology 1982 Apr;143(1):29-36. [Medline: 7063747]
28. Aronson AR. Effective mapping of biomedical text to the UMLS Metathesaurus: the MetaMap program. Proc AMIA Symp 2001:17-21. [Medline: 11825149]
29. Cimino JJ. Auditing the Unified Medical Language System with semantic methods. J Am Med Inform Assoc 1998 Jan;5(1):41-51 [FREE Full text] [Medline: 9452984]
30. Bodenreider O, Nelson SJ, Hole WT, Chang HF. Beyond synonymy: exploiting the UMLS semantics in mapping vocabularies. Proc AMIA Symp 1998:815-819. [Medline: 9929332]
31. Humphreys BL, Mccray AT, Cheh ML. Evaluating the coverage of controlled health data terminologies: report on the results of the NLM/AHCPR large scale vocabulary test. J Am Med Inform Assoc 1997 Nov;4(6):484-500 [FREE Full text] [Medline: 9391936]

Abbreviations

ATR: automated text recognition
AUC: area under the curve
CHV: consumer health vocabulary
NIH: National Institutes of Health
NLM: National Library of Medicine
OAC: open access and collaborative
POS: part of speech
ROC: receiver operating characteristic
UMLS: Unified Medical Language System

submitted 23.10.06; peer-reviewed by T Patrick; comments to author 21.11.06; revised version received 21.11.06; accepted 22.02.07; published 14.03.07.

Please cite as:
Zeng QT, Tse T, Divita G, Keselman A, Crowell J, Browne AC, Goryachev S, Ngo L
Term Identification Methods for Consumer Health Vocabulary Development
J Med Internet Res 2007;9(1):e4
URL: http://www.jmir.org/2007/1/e4/
doi:10.2196/jmir.9.1.e4
PMID:17478413

© Qing T Zeng, Tony Tse, Guy Divita, Alla Keselman, Jon Crowell, Allen C Browne, Sergey Goryachev, Long Ngo. Originally published in the Journal of Medical Internet Research (http://www.jmir.org), 14.03.2007. Except where otherwise noted, articles published in the Journal of Medical Internet Research are distributed under the terms of the Creative Commons Attribution License

XSL•FO
RenderX

(http://www.creativecommons.org/licenses/by/2.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited, including full bibliographic details and the URL (see "please cite as" above), and this statement is included.

XSL•FO
RenderX

Original Paper

Psychological Assessment via the Internet: A Reliability and Validity Study of Online (vs Paper-and-Pencil) Versions of the General Health Questionnaire-28 (GHQ-28) and the Symptoms Check-List-90-Revised (SCL-90-R)

Miguel A Vallejo[1], PhD; Carlos M Jordán[1], MS Psych; Marta I Díaz[1], PhD; María I Comeche[1], PhD; José Ortega[1], MS Psych

Facultad de Psicología, Universidad Nacional de Educación a Distancia (UNED), Madrid, Spain

Corresponding Author:
Miguel A Vallejo, PhD
Facultad de Psicología
Universidad Nacional de Educación a Distancia (UNED)
C/Juan del Rosal, s/n CP: 28040
Madrid, Spain
Phone: +34 91 398 79 36
Fax: +34 91 398 62 98
Email: mvallejo@psi.uned.es

Abstract

Background: Internet psychology services are rapidly increasing and that implies online assessment. To guarantee the results of these new online evaluation procedures, it is necessary to have reliable and valid assessment tools.

Objective: In this work we analyzed the online versions of two popular psychopathology screening questionnaires: the General Health Questionnaire-28 (GHQ-28) and the Symptoms Check-List-90-Revised (SCL-90-R).

Methods: A total of 185 psychology students were recruited from two universities in Madrid, Spain. All of them had Internet access at home. A test-retest situation and factorial analysis were used to generate reliability and validity data. Both paper-and-pencil questionnaires (test) and their online versions (retest) were completed by 100 participants (median gap = 17 days).

Results: Results suggest that both online questionnaires were fairly equivalent to their paper-and-pencil versions, with higher reliability values for the SCL-90-R. Factorial analysis tended to reproduce the structure shown in former investigations of both questionnaires, replicating the four-factor structure of the GHQ-28 but failing to do so with the nine-factor structure of the SCL-90-R. Instead, a large unrotated factor appeared.

Conclusions: Further research should be carried out to confirm these data, but our work supports the online use of both assessment tools. The psychometric properties of the online version of GHQ-28 is similar to the paper-and-pencil and we can recommend its utilization in a Web environment. In contrast, SCL-90-R can only be recommended as a global index for psychological distress, using the Global Severity Index (GSI), not necessarily its subscales; and it should be considered that the online scores were lower than the ones with the paper-and-pencil version.

(J Med Internet Res 2007;9(1):e2) doi:10.2196/jmir.9.1.e2

KEYWORDS

assessment; Internet; online

Introduction

Nowadays, online assessment is becoming necessary as clinical psychology is considering the Internet as a medium through which therapy and counseling can be offered [1]. It has already been shown how easy it is to create a website containing tools to assess psychological problems or constructs [2]. Moreover, the advantages over the traditional way of gathering data, such as easy and immediate scoring and missing data handling, have been made evident [3]. At this point, the reliability and validity of online questionnaires have become current and relevant research topics.

So far, the question “Will the mode of administration affect the respondent’s score?” has barely been formulated, and research on this topic has been undertaken by only a few studies.

XSL•FO
RenderX

Concepts related to social desirability [4], self-disclosure [5], or computer anxiety [6] are suggested as modulating variables that could modify the attitude toward computerized tests. Despite literature on these subjects, the research is still scarce and inconclusive and points to the need for further research to compare data from paper-and-pencil and online versions.

In that sense, a growing number of computerized or online questionnaires related to areas such as panic/agoraphobia [3], youth independence living [7], aggression and impulsivity [8], quality of life in diabetes [9], and a battery of 16 other health-related questionnaires [10] have already been studied. All but one of the computer/online versions (the Aggression Questionnaire by Buss and Perry [11]) were declared equivalent to their respective paper-and-pencil tests. Along with this, randomized studies on psychological distress tests have shown the same equivalence between the online and paper-and-pencil versions [12]. Nevertheless, in spite of the positive results supporting online assessment, the study of psychometric properties of online tests has frequent methodological problems (lack of random assignment or differing demographic characteristics to ensure sample equivalence), which make the adequate reliability or equivalence analysis difficult [13].

Taking the current state of the research into account, the present work aimed to obtain reliability and validity data for the online versions of two of the most frequently used psychopathology screening questionnaires in mental health: the General Health Questionnaire-28 (GHQ-28) [14] and the Symptoms Check-List-90-Revised (SCL-90-R) [15]. This paper is part of more extensive research aiming to develop a psychological treatment website following previous analysis of online clinical psychology websites in Spain [16]. Both questionnaires are used in a counseling website, the preliminary phase of the psychological treatment site. This choice was based on the wide research on and the historical use of these two questionnaires in psychopathology [17], as well as by their simple self-report structure, which makes it easy to incorporate them into a website.

A test-retest situation was chosen to obtain the reliability and validity data. Reliability was calculated as internal consistency, and test-retest correlation served as an equivalence index of the two test administration methods (paper-and-pencil and online). Inner structure exploration by factorial analysis was used to evaluate the construct validity of online versions. Although both questionnaires have a general score, they are divided into scales proposed as psychological disorders markers. The four scales of GHQ-28 (A: somatic symptoms, B: anxiety/insomnia, C: social dysfunction, and D: depression) have been found as a four-factor structure in previous studies [18-20]. For SCL-90-R, its nine scales (somatization, obsessive-compulsive symptoms, interpersonal sensitivity, depression, anxiety, hostility, phobic anxiety, paranoid ideation, psychoticism) were originally proposed as representing a nine-factor structure [21], but most of the research to date has failed to replicate this and has instead found either a primary global distress factor [22-26] or a four-, five-, or six-factor solution [26].

In short, with this work we try to contribute some of the needed empirical supporting data in order to ensure that online questionnaires have at least the psychometric characteristics attributed to their corresponding paper-and-pencil versions.

Methods

Sample

Participants were 185 psychology students recruited from two universities in Madrid, Spain. All of them had Internet access at home. This was a requirement to participate in the study in order to informally control how familiar participants were with the required technology. Although Internet familiarity is not a representative feature of the general population in Spain, this work is framed into a project in which the final point will be the development of a treatment website for mood disorders, so the sample resembles the target population in Internet familiarity.

Instruments

GHQ-28

The General Health Questionnaire (GHQ) is used to detect psychiatric disorder in the general population and within community or non-psychiatric clinical settings such as primary care or general medical outpatients. In the GHQ-28 the respondent is asked to compare his recent psychological state with his usual state. It is therefore sensitive to short-term psychiatric disorders but not to long-standing attributes of the respondent. All items have a 4 point scoring system using Likert scoring (0-1-2-3). The GHQ-28 contains 28 items that, through factor analysis, have been divided into four subscales, as mentioned above.

The Spanish-language version of the General Health Questionnaire by Lobo and Muñoz [28] was used. In the online version, one could scroll through the whole test. A pull-down menu in which the possible answers appeared followed the text of each item.

SCL-90-R

The Symptom Checklist-90-R (SCL-90-R) instrument has been designed to evaluate a broad range of psychological problems and symptoms of psychopathology. The instrument is also useful in measuring patient progress or treatment outcomes.

The SCL-90-R has 9 subscales, as mentioned above and in Table 4. The sum of all 9 subscales is the Global Severity Index (GSI), which can be used as a summary of the test, reflecting overall psychological distress.

We used the Spanish-language version of the Symptoms Check-List-90-Revised by González de Rivera et al [29]. The same online display method was used as for the GHQ-28.

Procedure

A classic test-retest design was carried out: the paper-and-pencil version of the instrument was used for the test and the online version for the retest. After verbally agreeing to participate in the study, participants received a booklet containing instructions, sociodemographic questions, and both screening questionnaires in paper-and pencil-format. At the end of the instructions page there was a box with the address of the website containing the

XSL•FO
RenderX

online questionnaires and the dates the site would be available. Identification of participants' online questionnaires was achieved by a nickname chosen and written by each subject in the questionnaire booklet.

To ensure that participants completed the online tasks, email addresses were requested in order to provide reminder messages (22 participants refused). Individual messages were sent 14 days after the paper-and-pencil task. A second reminder was sent if the online questionnaires were not received within a week after the first message.

Statistical Analysis

Statistical analysis was carried out with SPSS version 12.0. Reliability as internal consistency measured by Cronbach alpha was tested for both formats of the questionnaires and their subscales. Pearson correlation was used to prove the equivalence between paper-and-pencil forms and the online versions. A *t* test served to evaluate if there were statistically significant differences between the mean scores of the formats. We also applied η^2 after a repeated measures ANOVA. η2 is a measure of effect size in ANOVA: the degree of association between an effect (e.g., a main effect, an interaction, and a linear contrast) and the dependent variable. We used this statistic in trying to decide whether mean score differences have clinical relevance. Different benchmarks have been used to interpret η^2, but as for the $P < 0.05$ rule in hypothesis testing, there is only a rough guide to be used when no literature is available to compare effect size values, and the best way to interpret it must consider what outcome is being studied [30]. As this "rough guide," we will use $\eta^2 = .01 – .09$ for a small effect, $\eta^2 = .10 – .24$ for medium effects, and $\eta^2 \geq .25$ for large effects [31].

As stated earlier, construct validity was evaluated by means of principal components factorial analysis. Factorial structures similar to the ones shown in previous investigations following varimax rotation were expected, that is, four factors in GHQ-28 and nine in SCL-90-R. We also analyzed the unrotated solution and the sampling adequacy, using the Kaiser-Mayer-Olkin (KMO) test.

Results

Participant Demographics

From the initial sample of 185 participants, 104 completed both online questionnaires. This represents 56% of the total sample and 63% of those who received reminder messages. Although missing data was not possible online, four participants were rejected because of paper-and-pencil missing data, so 100 questionnaires were actually analyzed (Table 1). The majority of retests were received around the 14th day after the test (median = 17 days; min = 14, max = 38), and 90% had been received after 28 days.

XSL•FO
RenderX

Table 1. Demographic characteristics of the participants who completed all questionnaires, N = 100

Demographic characteristic	Number*
Age (years), mean (SD)	27.4 (10.01)
Number of children, mean (SD)	0.28 (0.72)
Gender	
Male	22
Female	78
Marital status	
Never married	83
Married	17
Education	
Technical education	3
High school	51
Short-term graduate	24
Long-term graduate	20
Undetermined	2
Socioeconomic status	
High	1
Middle-high	29
Middle	61
Middle-low	4
Low	3
Undetermined	2
Employment	
Full-time student	64
Worker/student	32
Unemployed/student	4

*Values are number unless otherwise noted.

GHQ-28

Reliability results for the GHQ-28 are shown in Table 2. Cronbach alpha was .90 for the whole test in both the paper-and-pencil and online formats, and it ranged from .71 to .85 among the scales, with scale C (social dysfunction) showing lower values in both formats.

Table 2. GHQ-28 reliability results: Cronbach alpha, mean differences t test, Pearson test-retest correlation, and squared eta effect size

	Paper-and-Pencil		Online				
GHQ-28	**Mean (SD)**	**α**	**Mean (SD)**	**α**	***t* (*P* value)**	***r* (*P* value)**	**η^2**
Scale A	4.77 (3.53)	.84	4.32 (3.48)	.84	−1.32 (.190)	.53 (< .001)	.017
Scale B	4.86 (3.80)	.83	4.19 (3.35)	.79	−2.45 (.016)	.72 (< .001)	.057
Scale C	6.81 (2.33)	.71	6.91 (2.21)	.79	.37 (.710)	.30 (.002)	.001
Scale D	1.27 (2.35)	.85	1.21 (2.51)	.84	−.30 (.769)	.65 (< .001)	.001
Total	17.71 (9.13)	.90	16.63 (9.00)	.90	−1.523 (.131)	.69 (< .001)	.023

Test-retest data showed significant correlations, ranging from .30 for scale C to .72 for scale B. Total score test-retest correlation was .69. We did a *t* test to see whether differences between scores from the two formats appeared. This occurred in scale B—paper-and-pencil scores were higher than online scores. We then used η^2 to check how big this difference was if taken as an effect size: its value was small (.057), being in

XSL•FO
RenderX

the same range as for those scales in which mean differences were not statistically significant (see Table 2).

The factorial analysis of GHQ-28 reproduces fairly well the presupposed four-factors solution in both the online and paper-and-pencil administrations. Table 3 represents item factorial loads among factors. Taking .30 or larger loads to assign each item to a factor, in both the online and paper-and-pencil analysis factor 1 includes all depression items (D), with the exception of online item D5, and factor 4 includes all social dysfunction (C) items, except paper-and-pencil item C2. Factor 2 grouped B (anxiety) items online and A (somatization) paper-and-pencil items. Factor 3 does the opposite, corresponding to A items online and B items in paper-and-pencil, except for B5. So, it could be said that each factor is close to its clinical interpretation. Nevertheless, a few items have bigger loads than expected in other factors. Scales A and B share large loads, a fact quite understandable given that somatization and anxiety appear together several times. Item D5 did not load at all in factor 1 in the online version, but did in factors 2 and 3. This could be explained by the meaning of the word “nerves” (included in the text of this item) identifying closer to anxiety than to depression. Item D5’s large load on paper-and-pencil factor 2 supports this interpretation. Lastly, paper-and-pencil scale C has smaller loads than expected in factor 4 in three of its seven items. We will interpret this alongside scale C’s test-retest correlation later.

The predominantly positive values in the original correlation matrixes suggest paying attention to a general unrotated factor that could explain some of the item sharing among scales. This general factor explained 28.44% (paper-and-pencil) and 29.48% (online) of the variance, and 27 and 26 (paper-and-pencil and online, respectively) out of 28 items had loads of .30 or greater.

XSL•FO
RenderX

Table 3. GHQ-28 item factorial loads (varimax rotation): A: somatization, B: anxiety; C: social dysfunction; D: depression

	Paper-and-Pencil (53.98%)*				Online (55.85%)			
Item	**1 (16.18)†**	**2 (14.54)**	**3 (13.59)**	**4 (09.68)**	**1 (16.07)**	**2 (14.33)**	**3 (13.06)**	**4 (12.42)**
A1	.069	**.607‡**	.007	**.313**	.000	**.327**	**.416**	**.449**
A2	-.056	**.687**	**.328**	.282	.025	**.502**	**.486**	.254
A3	.121	**.759**	**.361**	.119	.180	**.591**	**.460**	.251
A4	.166	**.771**	-.066	.191	.290	.237	**.613**	.215
A5	.040	**.574**	.294	-.143	.097	.127	**.657**	.029
A6	.128	**.669**	.197	-.232	.045	.076	**.744**	.033
A7	.181	**.603**	.068	-.066	.236	**.326**	**.471**	.036
B1	.056	**.332**	**.575**	.153	**.303**	**.653**	-.071	-.036
B2	.029	**.317**	**.360**	.270	-.001	**.391**	.126	-.050
B3	.253	.218	**.796**	.017	-.040	**.722**	.207	.110
B4	.212	.067	**.815**	.090	.130	**.703**	.248	.240
B5	**.390**	**.363**	.226	-.141	.000	**.322**	**.547**	-.004
B6	.210	.078	**.775**	.113	.343	**.490**	**.315**	.179
B7	.182	.176	**.716**	.032	.111	**.719**	.271	.120
C1	-.055	-.084	-.224	**.501**	.213.	**-.337**	.195	**.578**
C2	**.456**	.291	.206	.213	**.422**	.074	-.086	**.537**
C3	.105	.042	.118	**.786**	-.007	.136	.140	**.682**
C4	.066	.154	.175	**.780**	.224	-.049	.263	**.711**
C5	.016	.065	.253	**.666**	.115	.186	-.214	**.675**
C6	**.366**	-.098	.139	**.312**	.051	.099	.048	**.710**
C7	**.317**	.265	**.345**	**.315**	.154	**.484**	-.051	**.553**
D1	**.406**	.222	.239	.032	**.463**	.051	**.567**	-.066
D2	**.779**	.226	.117	.076	**.889**	.127	.191	.210
D3	**.682**	.124	.073	.265	**.873**	.137	.073	.129
D4	**.868**	-.022	.108	-.009	**.842**	-.050	.261	.105
D5	**.491**	**.403**	.240	-.128	-.046	**.416**	**.448**	.037
D6	**.832**	.089	.096	-.029	**.872**	.133	.182	.058
D7	**.838**	-.008	.129	-.076	**.724**	.120	-.081	.197

*The percentage of the variance explained by the whole model appears in parentheses.

†The percentage of the variance explained by each factor appears in parentheses.

‡Bold type identifies .30 or larger loads.

SCL-90-R

Table 4 shows reliability data for the SCL-90-R. The Cronbach alpha of the global severity index (GSI) was .96 and .97 for the paper-and-pencil and online versions, respectively. Scales showed .72 or higher except for phobic anxiety in the paper-and-pencil questionnaire, which was .62. Test-retest correlation ranged from .63 for hostility to .86 for psychoticism. The correlation for the GSI was .83.

XSL•FO
RenderX

Table 4. SCL-90-R reliability results: Cronbach alpha, mean differences t test, Pearson test-retest correlation, and squared eta effect size

	Paper-and-Pencil		Online				
SCL-90-R	**Mean (SD)**	α	**Mean (SD)**	α	***t* (*P* value)**	***r* (*P* value)**	η^2
Somatization	.58 (.55)	.85	.39 (.41)	.82	5.10 (< .001)	.73 (< .001)	.208
Obsessive-Compulsive	.61 (.59)	.84	.45 (.50)	.85	4.10 (< .001)	.74 (< .001)	.145
Interpersonal sensitivity	.62 (.51)	.78	.45 (.51)	.83	5.54 (< .001)	.83 (< .001)	.236
Depression	.66 (60)	.88	.53 (.62)	.92	2.91 (.004)	.74 (< .001)	.079
Anxiety	.47 (.46)	.81	.36 (.38)	.79	3.30 (.001)	.70 (< .001)	.099
Hostility	.40 (.43)	.72	.29 (.41)	.76	3.01 (.003)	.63 (< .001)	.084
Phobic anxiety	.18 (.31)	.62	.15 (.29)	.72	1.56 (.123)	.84 (< .001)	.024
Paranoid ideation	.42 (.51)	.73	.34 (.49)	.81	2.41 (.017)	.77 (< .001)	.056
Psychoticism	.20 (.33)	.75	.19 (.32)	.76	.812 (.419)	.86 (< .001)	.007
GSI (Total)	.50 (.40)	.96	.37 (.36)	.97	5.47 (< .001)	.83 (< .001)	.232

Paper-and-pencil means were higher than online means in every score. A *t* test for repeated measures showed that those differences were statistically significant except for phobic anxiety and psychoticism. Squared eta (η^2) analysis showed values from small to medium. It is important to note that η^2 for the GSI was .232, which means that more than 23% of the variance was due to method administration (see Table 4). That proportion could have clinical implications that we will discuss later.

The factorial analysis showed difficulty confirming the expected nine-factors solution for both the online and paper-and-pencil administration. All the items where scattered through the forced nine factors without the presupposed order. As an example, we could mention that the first online factor grouped items (.30 or bigger loads) from seven theoretical scales (anxiety, hostility, depression, interpersonal sensitivity, obsessive-compulsive symptoms, phobic anxiety, and psychoticism). Another fact lead us to reject a factorial analysis for this questionnaire: the KMO test (online = .394; paper-and-pencil = .414) was under the recommended .6 value to accept such an analysis [32]. As a comparison, the GHQ-28 KMO values were .788 for the online version and .781 for the paper-and-pencil version. As a result we do not recommend the use of the SCL-90-R scales as the way to discriminate among different clinical problems.

However, it should be noted that the first unrotated component of the analysis explained more than 25% of the variance in both online and paper-and-pencil questionnaires, and 94% of the online items (85 out 90) and 92% of the paper-and-pencil ones (83 out 90) presented loads of .30 or higher for this general factor. This, together with reliability data, led us to accept this test as a general screening tool.

Discussion

The aim of this work was to find out whether the psychometric characteristics of two well-known, self-report questionnaires remain consistent when administered via the Internet. Our analysis of the online versions matches the results of the paper-and-pencil versions in several aspects, but some identified differences between the two formats should be explained.

GHQ-28

Regarding the GHQ-28, internal consistency was high in both formats (Cronbach alpha for all scales and total score was over .70). Nevertheless, test-retest reliability ranged from a too modest .30 to .72, while other studies have presented coefficients over .70, some of them using Spanish translations of the questionnaire [33]. On one hand, it could be said that the GHQ-28 keeps its reliability as internal consistency when delivered via the Internet, but, on the other hand, equivalence data are lower than expected, especially in scale C. The small test-retest correlation in this scale (.30) as well as its factorial instability in the paper-and-pencil version could be due to the experimental situation. C scale accounts for "social dysfunction," and the paper-and-pencil situation was "social" (all the students and the investigator were together in the same classroom), whereas the online task was completed at home. Perhaps this caused participants to interpret the C items differently and to vary their answers.

Mean differences between formats were small enough to be negligible if we take into account η^2 results. Even in scale B, where these differences were significant, the accounted variance for method administration was only 5.7%, a proportion not very important when talking about a rough general screening test.

Validity analysis of GHQ-28 showed that previously reported factor structure was fairly replicated. As a whole, both online and paper-and-pencil results of this study match former works in which scales C (social dysfunction) and D (depression) were more consistent than A (somatic symptoms) and B (anxiety) [33,34]. This situation is clinically understandable given that somatic symptoms are frequent in anxiety disorders. A tentative explanation for the relative instability of the online C factor based on the experimental situation has already been pointed out.

SCL-90-R

The SCL-90-R maintained its internal consistency when delivered over the Internet; in fact, it was higher than in the

XSL•FO
RenderX

paper-and-pencil version, and test-retest correlations were as high as in previous studies [26]. This leads us to propose equivalence of the online and paper-and-pencil formats. Our results match the literature on reliability as internal consistency in nonclinical samples [26] as well as the equivalence data using an SCL-90-R computerized version [17]. Nevertheless, all paper-and-pencil scores were higher than online ones. Here it is important to mention the η^2 values. Three scale differences and that for the GSI could be labeled as medium effects. As we mentioned above, in the case of GSI, this means that 23.2% of the variance could be explained by test administration method. This proportion is big enough to recommend caution if we mixed online and traditional versions of this test because scores could differ enough as to cover (if online is first) or to resemble (if paper is first) the effect of a treatment. The presence of the experimenter and the participants during the paper-and-pencil session, plus the fact that all participants had Internet connections at home, leads us to believe that the online tasks were less aversive. This could be a tentative explanation of higher paper-and-pencil scores.

We have already mentioned the problems that most authors have faced when replicating the nine-factor structure of the SCL-90-R. In our case, the more parsimonious interpretation matches the conclusions of several articles: even when the proposed solution has more than a factor [22], the high variance percentage explained by the first factor should lead to consideration of the total score as a general dimension of psychopathology [26]. Perhaps, as stated by Cyr et al [23], “interpreting nine dimensions for clinical purposes is highly questionable” no matter if we are talking about online or classic assessment. As only one strong factor appears, a psychopathology discrimination function can not be assigned to this tool. However, it does not lose its usefulness as a general psychopathological screening tool.

Conclusion

The results of this research are encouraging for the online use of the two questionnaires. In the GHQ-28, although two of its four scales had relatively small equivalence values, those of the other two as well as that of the general score were adequate, and the internal consistency values were high. Further research should be carried out to confirm this data, but our work supports the online use of this assessment tool.

The same could be said about the SCL-90-R: its online version could be taken as being equivalent to its classic paper-and-pencil version, and its internal consistency is high. However, paper-and-pencil scores were higher than online ones. Even when an online test has shown acceptable reliability and validity values, the use of normative data from paper-and-pencil questionnaires may not be appropriate [2], suggesting that as online testing spreads, research to obtain a bank of normative data from larger Internet samples should be an important goal.

Factorial analysis results for both online questionnaires showed factor structures similar to paper-and-pencil versions. SCL-90-R showed a similar factorial structure in its online and paper-and-pencil applications, but the results do not replicate the nine factor structure proposed by Derogatis [21]. Other researchers also had difficulties to replicate the nine factors [21,23,26]. As a consequence, we recommend use of the questionnaire as a general index of psychopathology, using the summary score (GSI) only, not the subscales.

The use of standardized tools administered through the Internet needs further investigation, and as for paper-and-pencil versions, they are not enough to properly assess a clinical case. The results obtained by these screening tools should be taken only as part of the assessment and should never be used as the only basis to support any intervention.

Lastly, we should mention two limitations of this work that future research should try to address. First, as the most probable Internet users, the university community will be one of the target populations for any Internet-related research. We must stress that this technology is spreading fast, so samples outside the university community must be analyzed. Second, our experimental design did not allow us to separate the effects of the test-retest situation from those of the format effect. Therefore, the next step should be to compare four groups (Internet and Internet; Internet and paper-and-pencil; paper-and-pencil and paper-and-pencil; paper-and-pencil and Internet) to discriminate both effects.

Acknowledgments

This work was carried out within the framework of the Research Grant MEC I+D+I SEJ2004-03392/PSI, director Miguel A Vallejo. We thank María Crespo, Cristina Larroy, and Begoña Rosales for their help obtaining the participation of the subjects for this research.

Conflicts of Interest

None declared.

Multimedia Appendix

Psychological assessment through the Internet [PPT file (MS PowerPoint), 328 KB - jmir_v9i1e2_app1.ppt]

References

1. Carlbring P, Nilsson-Ihrfelt E, Waara J, Kollenstam C, Buhrman M, Kaldo V, et al. Treatment of panic disorder: live therapy vs. self-help via the Internet 2005 Oct;43(10):1321-1333. [Medline: 16086983]

XSL•FO
RenderX

2. Buchanan T. Internet-based questionnaire assessment: appropriate use in clinical contexts. Cogn Behav Ther 2003;32(3):100-109. [Medline: 16291542] [doi: 10.1080/16506070310000957]
3. Carlbring P, Brunt S, Bohman S, Richards P, Öst LG, Andersson G. Internet vs. paper and pencil administration of questionnaires commonly used in panic/agoraphobia research. Comput Human Behav 2007;23(3):1421 . In press. [doi: 10.1016/j.chb.2005.05.002]
4. Booth-Kewley S, Larson GE, Miyoshi DK. Social desirability effects on computerized and paper-and-pencil questionnaires. Comput Human Behav 2007;23(1) . In press. [doi: 10.1016/j.chb.2004.10.020]
5. Suler J. The online disinhibition effect. Cyberpsychol Behav 2004 Jun;7(3):321-326. [Medline: 15257832] [doi: 10.1089/1094931041291295]
6. Schulenberg SE, Yutrzenka BA. Ethical issues in the use of computerized assessment. Comput Human Behav 2004 Jul;20(4):477-490. [doi: 10.1016/j.chb.2003.10.006]
7. Bressani RV, Downs AC. Youth independent living assessment: testing the equivalence of web and paper/pencil versions of the Ansell-Casey Life Skills Assessment. Comput Human Behav 2002 Jul;18(4):453-464. [doi: 10.1016/S0747-5632(01)00053-X]
8. Surís A, Borman PD, Lind L, Kashner TM. Aggression, impulsivity, and health functioning in a veteran population: equivalency and test-retest reliability of computerised and paper-and-pencil administrations. Comput Human Behav . In press. [doi: 10.1016/j.chb.2004.03.038]
9. Pouwer F, Snoek FJ, Van Der Ploeg HM, Heine RJ, Brand AN. A comparison of the standard and the computerized versions of the Well-being Questionnaire (WBQ) and the Diabetes Treatment Satisfaction Questionnaire (DTSQ). Qual Life Res 1998 Jan;7(1):33-38. [Medline: 9481149] [doi: 10.1023/A:1008832821181]
10. Ritter P, Lorig K, Laurent D, Matthews K. Internet versus mailed questionnaires: a randomized comparison. J Med Internet Res 2004 Sep 15;6(3):e29 [FREE Full text] [Medline: 15471755] [doi: 10.2196/jmir.6.3.e29]
11. Buss AH, Perry M. The aggression questionnaire. J Pers Soc Psychol 1992 Sep;63(3):452-459. [Medline: 1403624] [doi: 10.1037/0022-3514.63.3.452]
12. Herrero J, Menese J. Short Web-based versions of the perceived stress (PSS) and Center for Epidemiological Studies-Depression (CESD) scales: a comparison to pencil and paper responses among Internet users. Comput Human Behav 2006 Sep;22(5):830-846. [doi: 10.1016/j.chb.2004.03.007]
13. Epstein J, Klinkenberg WD, Wiley D, McKinley L. Insuring sample equivalence across Internet and paper-and-pencil assessments. Comput Human Behav 2001 May;17(3):339-346. [doi: 10.1016/S0747-5632(01)00002-4]
14. Goldberg DP, Hillier VF. A scaled version of the General Health Questionnaire. Psychol Med 1979 Feb;9(1):139-145. [Medline: 424481]
15. Derogatis LR. SCL-90-R: Administration, scoring and procedures manual for the R(evised) version. Baltimore, MD: John Hopkins University, School of Medicine. 1977.
16. Vallejo MA, Jordán CM, Mañanes G, Vasallo A, Comeche MI. Clinical psychology offers in the Internet in Spain. Comput Human Behav . In press. [doi: 10.1016/j.chb.2006.02.016]
17. Schmitz N, Kruse J, Heckrath C, Alberti L, Tress W. Diagnosing mental disorders in primary care: the General Health Questionnaire (GHQ) and the Symptom Check List (SCL-90-R) as screening instruments. Soc Psychiatry Psychiatr Epidemiol 1999 Jul;34(7):360-366. [Medline: 10477956] [doi: 10.1007/s001270050156]
18. Aderibigbe YA, Riley W, Lewin T, Gureje O. Factor structure of the 28-item general health questionnaire in a sample of antenatal women. Int J Psychiatry Med 1996;26(3):263-269. [Medline: 8976467]
19. Iwata N, Saito K. The factor structure of the 28-item General Health Questionnaire when used in Japanese early adolescents and adult employees: age- and cross-cultural comparisons. Eur Arch Psychiatry Clin Neurosci 1992;242(2-3):172-178. [Medline: 1486106] [doi: 10.1007/BF02191565]
20. Kilic C, Rezaki M, Rezaki B, Kaplan I, Ozgen G, Sagduyu A. Ozturk MO General Health Questionnaire (GHQ-12 & GHQ-28): psychometric properties and factor structure of the scales in a Turkish primary care sample. Soc Psychiatry Psychiatr Epidemiol 1997 Aug;32(6):327-331. [Medline: 9299926]
21. Derogatis LR, Cleary PA. Confirmation of the dimensional structure of the SCL-90: a study in construct validation. J Clin Psychol 1977 Oct;33(4):981-989.
22. Brophy CJ, Norvell NK, Kiluk DJ. An examination of the factor structure and convergent and discriminant validity of the SCL-90R in an outpatient clinic population. J Pers Assess 1988 Dec;52(2):334-340. [Medline: 3404394] [doi: 10.1207/s15327752jpa5202_14]
23. Cyr JJ, McKenna-Foley JM, Peacock E. Factor structure of the SCL-90-R: is there one? J Pers Assess 1985 Dec;49(6):571-578. [Medline: 4093836] [doi: 10.1207/s15327752jpa4906_2]
24. Martínez-Azumendi O, Fernández-Gómez C, Beitia-Fernández M. Factorial variance of the SCL-90-R in a Spanish out-patient psychiatric sample. Actas Esp Psiquiatr 2001 Mar;29(2):95-102. [Medline: 11333527]
25. Rauter UK, Leonard CE, Swett CP. SCL-90-R factor structure in an acute, involuntary, adult psychiatric inpatient sample. J Clin Psychol 1996 Nov;52(6):625-629. [Medline: 8912105] [doi: 10.1002/(SICI)1097-4679(199611)52:6<625::AID-JCLP4>3.0.CO;2-J]

XSL•FO
RenderX

26. Schmitz N, Hartkamp N, Kiuse J, Franke GH, Reister G, Tress W. The Symptom Check-List-90-R (SCL-90-R): a German validation study. Qual Life Res 2000 Mar;9(2):185-193. [Medline: 10983482] [doi: 10.1023/A:1008931926181]
27. ; INE (Instituto Nacional de Estadística). Estadística de la enseñanza universitaria en España. URL: http://www.ine.es/inebase/cgi/um?M=%2Ft13%2Fp405%2Fa2003-2004&O=pcaxis&N=&L=0 [accessed 2006 Apr 25] [WebCite Cache ID 5FPSUE1DA]
28. Lobo A, Muñoz PE. Cuestionario de salud general. GHQ (General Health Questionnaire). Guía para el usuario de las distintas versiones. Versiones en lengua española validadas. Barcelona, Spain: Masson; 1996.
29. González DRJL, Derogatis LR, De LCC, Garcia-Marco R, Rodríguez F, Henry R, et al. The Spanish version of the SCL-90-R. Normative data in the general population. Towson, MD: Clinical Psychometric Research; 1989.
30. Vacha-Haase T, Thompson B. How to estimate and interpret various effect sizes. J Couns Psychol 2004 Oct;51(4):473-481. [doi: 10.1037/0022-0167.51.4.473]
31. Cohen J. Multiple regression as a general data-analytic system. Psychol Bull 1968 Dec;70(6):426-443. [doi: 10.1037/h0026714]
32. Pardo A, Ruiz MA. SPSS 11.Guía para el análisis de datos. Madrid, Spain: McGraw-Hill; 2002.
33. Gibbons PF, Hilda M. Assessment of the factor structure and reliability of the 28-item version of the General Health Questionnaire (GHQ-28) in El Salvador. Int J Clin Health Psychol 2004 May;4(2):389-398 [FREE Full text]
34. Werneke U, Goldberg DP, Yalcin I, Ustün BT. The stability of the factor structure of the General Health Questionnaire. Psychol Med 2000 Jul;30(4):823-829. [Medline: 11037090] [doi: 10.1017/S0033291799002287]

Abbreviations

GHQ-28: General Health Questionnaire-28
GSI: global severity index
SCL-90-R: Symptoms Check-List-90 Revised

submitted 09.03.06; peer-reviewed by P Carlbring, U Werneke; comments to author 22.04.06; revised version received 27.10.06; accepted 28.12.06; published 31.01.07.

Please cite as:
Vallejo MA, Jordán CM, Díaz MI, Comeche MI, Ortega J
Psychological Assessment via the Internet: A Reliability and Validity Study of Online (vs Paper-and-Pencil) Versions of the General Health Questionnaire-28 (GHQ-28) and the Symptoms Check-List-90-Revised (SCL-90-R)
J Med Internet Res 2007;9(1):e2
URL: http://www.jmir.org/2007/1/e2/
doi:10.2196/jmir.9.1.e2
PMID:17478411

© Miguel A Vallejo, Carlos M Jordán, Marta I Díaz, María I Comeche, José Ortega. Originally published in the Journal of Medical Internet Research (http://www.jmir.org), 31.01.2007. Except where otherwise noted, articles published in the Journal of Medical Internet Research are distributed under the terms of the Creative Commons Attribution License (http://www.creativecommons.org/licenses/by/2.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited, including full bibliographic details and the URL (see "please cite as" above), and this statement is included.

XSL•FO
RenderX

Original Paper

Assessing Consumer Health Vocabulary Familiarity: An Exploratory Study

Alla Keselman[1,2], PhD, MA; Tony Tse[1], PhD; Jon Crowell[3], MS; Allen Browne[1], PhD; Long Ngo[3], PhD; Qing Zeng[3], PhD

[1]National Library of Medicine, National Institute of Health, Bethesda, MD, USA

[2]Aquilent, Inc, Laurel, MD, USA

[3]Decision Science Group, Brigham and Women's Hospital, Harvard Medical School, Boston, MA, USA

Corresponding Author:
Qing Zeng, PhD
Harvard Medical School
Decision Systems Group
Brigham and Women's Hospital
Thorn 304, 75 Francis Street
Boston, MA 02115
USA
Phone: +1 617 732 7694
Fax: +1 617 739 3672
Email: qzeng@dsg.harvard.edu

Abstract

Background: Accurate assessment of the difficulty of consumer health texts is a prerequisite for improving readability. General purpose readability formulas based primarily on word length are not well suited for the health domain, where short technical terms may be unfamiliar to consumers. To address this need, we previously developed a regression model for predicting "average familiarity" with consumer health vocabulary (CHV) terms.

Objective: The primary goal was to evaluate the ability of the CHV term familiarity model to predict (1) surface-level familiarity of health-related terms and (2) understanding of the underlying meaning (concept familiarity) among actual consumers. Secondary goals involved exploring the effect of demographic factors (eg, health literacy) on surface-level and concept-level familiarity and describing the relationship between the two levels of familiarity.

Methods: Survey instruments for assessing surface-level familiarity (45 items) and concept-level familiarity (15 items) were developed. All participants also completed a demographic survey and a standardized health literacy assessment, S-TOFHLA.

Results: Based on surveys completed by 52 consumers, linear regression suggests that predicted CHV term familiarity is a statistically significantly predictor ($P < .001$) of participants' surface-level and concept-level familiarity performance. Health literacy was a statistically significant predictor of surface-level familiarity scores ($P < .001$); its effect on concept-level familiarity scores warrants further investigation ($P = 0.06$). Educational level was not a significant predictor of either type of familiarity. Participant scores indicated that conceptualization lagged behind recognition, especially for terms predicted as "likely to be familiar" ($P = .006$).

Conclusions: This exploratory study suggests that the CHV term familiarity model is predictive of consumer recognition and understanding of terms in the health domain. Potential uses of such a model include readability formulas tailored to the consumer health domain and tools to "translate" professional medical documents into text that is more accessible to consumers. The study also highlights the usefulness of distinguishing between surface-level term familiarity and deeper concept understanding and presents one method for assessing familiarity at each level.

(J Med Internet Res 2007;9(1):e5) doi:10.2196/jmir.9.1.e5

KEYWORDS

Consumer health vocabulary; Patients; Vocabulary; Consumer health informatics; Health education; Readability; Comprehension; Health; Evaluation studies

XSL•FO
RenderX

Introduction

Improving the readability of online consumer health materials is an important area of eHealth research. Studies indicate that health information on the Web is beyond the reading ability of average consumers [1,2]. Research on general literacy suggests that readability decreases as the number of "difficult" words, those unfamiliar to the average reader, increases. Since familiarity correlates with education and literacy levels, "easy" terms are those that are familiar to many individuals who have lower reading skills. For example, the Dale-Chall readability formula incorporates a list of 3000 words and phrases (expressions) familiar to 80% of fourth-grade students in the United States [3]. However, because obtaining a comprehensive, empirically derived list of familiar words is difficult, many other existing readability formulas use average number of syllables per word as a surrogate for word difficulty.

Many researchers point to the need to reduce the gap between health literacy of the readers and the readability of consumer health materials [4]. As guidelines call for using simple common words, adhering to them requires predicting consumer familiarity with various health-related words. Currently, the only available methods are general purpose readability formulas developed by K-12 researchers. However, using such readability formulas to predict readers' ability to comprehend health texts has been criticized by the health literacy community. As McCray observes, "counting words and syllables and consulting a grade-level word list are most likely not sufficient to determine how readable a text is" [5]. Reliance on word length is particularly ill suited for the health domain, where short technical terms are likely to be unfamiliar to consumers (eg, apnea). The logic of graded word lists simplifies the phenomenon of word knowledge by implying that it is binary in nature and suggests that a reader is either unfamiliar or familiar with a particular word, with the switch between not knowing and knowing occurring at a single point in time. However, consumer health term familiarity is a more nuanced phenomenon involving partial knowledge [6], and increased exposure likely results in increased familiarity.

Recognizing the limitations of these previous approaches, we set out to explore alternative measures that account for "average" familiarity with health terms among members of a convenience sample of consumers. The ability to recognize terms is important because readers need to associate health terms with their corresponding concepts in order to extract useful information from text. Thus, we decompose health vocabulary knowledge into two parts: (1) surface-level term familiarity, or recognition of the lexical form, and (2) concept-level term familiarity, or understanding of the underlying concept. In cognitive science, a concept can be viewed as a set of slots that can be filled with characteristics describing a class of objects or events [7]. For instance, a "disease" concept may be characterized by attributes such as cause, severity, duration, and pathophysiology (among others). The completeness and accuracy of conceptual knowledge exists on a continuum, dictated by context. Thus, a healthy individual with a family history of diabetes and a diabetic patient may each benefit from explanations focusing on different aspects of diabetes (eg, prevention versus treatment). Yet, historically, readability studies do not distinguish between surface-level lexical forms (commonly referred to as "terms") and concepts and, therefore, do not separately assess familiarity at each "level."

We had previously developed a support vector machine regression model for predicting "familiarity likelihood scores" of consumer health vocabulary (CHV) terms using the empirical data from user studies evaluating "consumer-friendly display" names for medical concepts [8] as training data and the term frequency counts from health text corpora as features [9]. The model evaluated by this current study was an improved version of the initial model published in 2005 [9]: actual familiarity data were collected from 41 subjects for training, and term and word frequencies in three different corpora were used as features, including (1) Reuters news reports (health and non-health articles), (2) queries to a health search engine (MedlinePlus), and (3) queries to a general search engine (MetaCrawler). This algorithm assigns each consumer health term with a predictive score ranging from 0 to 1.0, representing the likelihood that a term is familiar to the average consumer. Terms are classified into three familiarity categories based on their scores: "likely" (> 0.8), "somewhat likely" (0.8-0.5), and "not likely" to be familiar (scores < 0.5).

The primary goal of the research reported in this paper was to develop and apply a simple methodology for validating the CHV familiarity predictive model against actual empirically derived familiarity with various health terms among health consumers. The validation is distinct and independent from the empirical data used in deriving the model. Both surface-level (ie, recognition) and concept-level familiarity (ie, understanding of the underlying meaning) data were collected from participants. Surface-level familiarity was investigated because it corresponds with existing conventional approaches to assessing health vocabulary knowledge. The goal of concept-level familiarity assessment was to explore the potential of this novel approach and to characterize the relationship between the two familiarity levels. Finally, we sought to describe the effect of demographic factors (including health literacy and education level) on actual consumers' scores. The following three hypotheses addressed the goals of the study:

1. Predicted familiarity likelihood level will have a significant effect on consumer surface-level term familiarity and consumer understanding of the underlying concept.
2. Demographic factors, including but not limited to health and education level, will have a significant effect on both types of familiarity scores.
3. Consumers' surface-level familiarity with terms will be greater than their understanding of the underlying concepts.

Methods

Participants

Consumers (n = 52) were recruited from Brigham and Women's Hospital. Health literacy, assessed with Short Test of Functional Health Literacy in Adults (S-TOFHLA) [10], ranged in score from 22 to 36 (mean = 33.04; SD = 3.83). Based on these scores, 50 participants had adequate health literacy skills (scores from

XSL•FO
RenderX

23-36 out of 36), while two had marginal skills (scores from 17-22).

Other demographic variables were self-reported using a brief questionnaire (Table 1). There were 8 non-native English speakers, with number of years speaking English ranging from 6 to 40 (median = 12 years). The level of English proficiency was not assessed, as the complexity of the relationship between primary and secondary language health literacy is beyond the scope of this study. Of the 8 non-native English participants, 7 achieved S-TOFHLA scores in the high literacy range, and the remaining participant, in the moderate literacy range (self-reported as speaking English for 40 years).

Table 1. Demographic characteristics of the participants (n = 52)

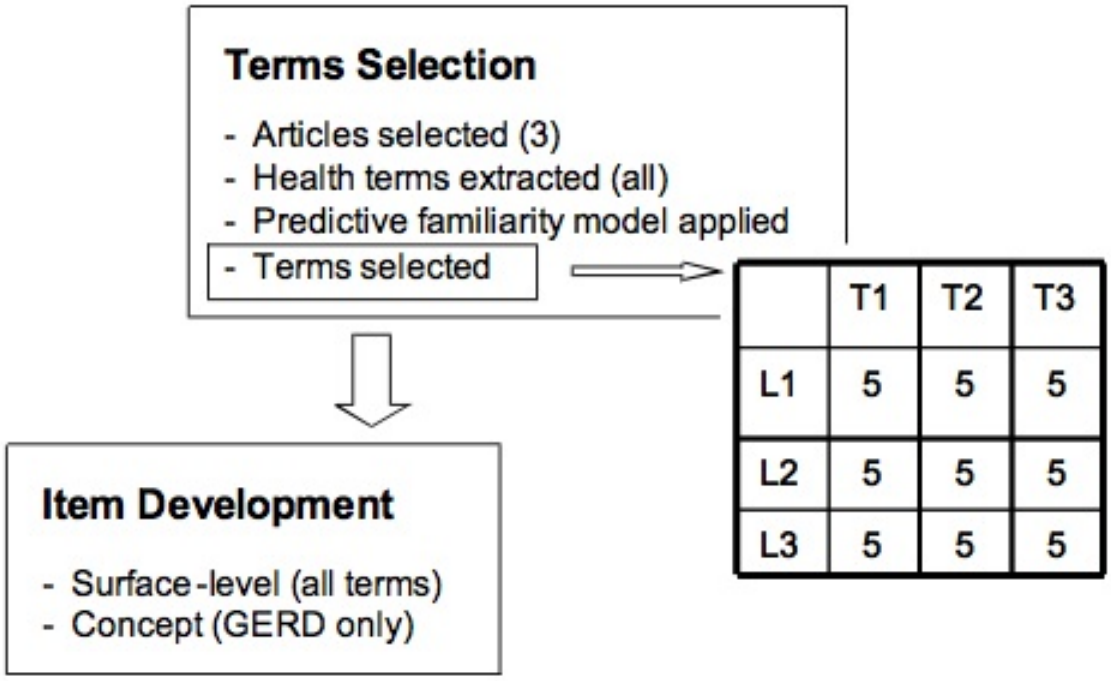
Demographic Variable	Number
Gender	
Male	16
Female	36
English proficiency	
Native speakers	44
Non-native speakers	8
Highest education level	
Below high school	2
High school	9
Some college	20
College	13
Graduate school	8
Age	
18-25	5
26-39	13
40-59	25
≥ 60	9
Race	
White	25
Black	13
Hispanic	8
Other	6
Health literacy level (STOFHLA scores)	
high health literacy (23-36)	50
moderate health literacy (17-22)	2

Instrument

A survey for assessing CHV surface-level (45 items) and concept-level (15 items) familiarity was developed, piloted tested, and implemented as described below. The process of instrument development consisted of two stages: (1) selecting health terms for inclusion in the test and (2) developing multiple-choice items for each term (Figure 1).

XSL•FO
RenderX

Figure 1. Survey development process (T = topic; L = predicted familiarity level)



Candidate CHV terms were selected from consumer health texts for three frequently visited MedlinePlus health topics: hypertension, back pain, and gastroesophageal reflux disease (GERD). One representative article on each selected topic was chosen from among consumer health sites listed by MedlinePlus. A final-year medical student manually extracted all health-related terms from each article. Next, all extracted terms were submitted to the predictive familiarity model [9] and assigned to the categories of "likely," "somewhat likely," or "unlikely" to be familiar. Finally, five terms from each predicted familiarity likelihood level were randomly selected from each of the three articles (Multimedia Appendix).

The next stage of instrument construction involved developing multiple-choice test items assessing the two types of familiarity, operationally defined as the following:

1. Surface-level familiarity: ability to match written health terms with basic relevant associated terms at the super-category, location, or function level (eg, "biopsy" is a "test")

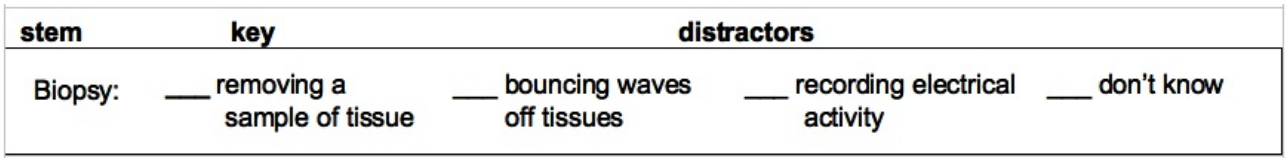
2. Concept-level familiarity: ability to associate written terms with brief phrases describing the meaning or "gists" (eg, "biopsy" means "removing a sample of tissue")

Surface-level familiarity items (Figure 2) were developed for all selected terms. Concept-level familiarity items (Figure 3) were developed only for the terms extracted from the article on GERD, in order to minimize survey administration time.

The layout of all test items was modeled on the Short Assessment of Health Literacy for Spanish-Speaking Adults (SAHLSA) [11], which in turn is based on the Rapid Estimate of Adult Literacy in Medicine (REALM) health literacy test for English speakers [12]. We chose the SAHLSA model because this validated instrument assesses the ability to associate health terms with other related terms. In contrast, REALM, commonly used in studies with English speakers, only tests the ability to pronounce health terms correctly, which we felt was less appropriate for our purpose of word knowledge assessment. SAHLSA is easy to administer and consists of 50 items, each with a "stem" or target term, a "key" term meaningfully associated with the target term, a "distractor," and a "don't know" option. Our only change to this format was adding a second "distractor" to reduce the probability of selecting the "key" term by guessing (see Figure 2). In developing CHV assessment items, we followed the following criteria: (1) the key term and distractors should be of the same difficulty as the target term, (2) distractors should be incorrect but plausible, and (3) the key term and distractors should have the same semantic relationship to the target term (eg, all location or all function). Criteria 2 and 3 were adopted from SAHLSA.

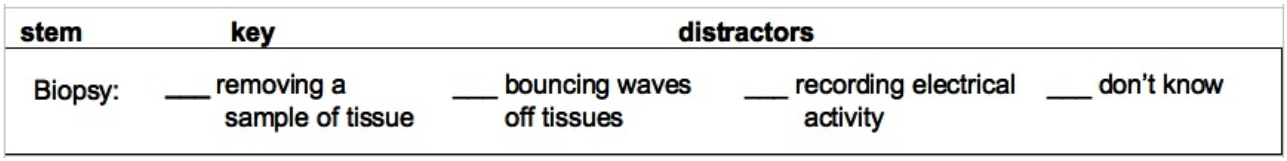
XSL•FO
RenderX

Figure 2. Sample CHV instrument surface-level familiarity item



Incorporating the REALM procedure, SAHLSA requires the examinee both to correctly pronounce the target term and to select the key term. However, since our goal was to measure familiarity with written health expressions and concepts explicitly using a self-administered tool (eg, via the Web), the SAHLSA requirement for examinees to pronounce each target expression was dropped. The final test included surface-level familiarity items for all three health topics (questions 1-45) and concept-level familiarity items for GERD terms only (questions 46-60). The entire instrument is available in the Multimedia Appendix.

Figure 3. Sample CHV instrument concept-level familiarity item



Administration, Scoring, and Analysis

Participants first completed the demographics survey, followed by the S-TOFHLA and CHV familiarity survey (surface-level items followed by concept-level familiarity items). For scoring, each correct answer was awarded one point. Surface-level and concept-level familiarity scores were calculated separately. Regression analysis tests on the data were performed at the 0.05 level of significance. Since the study is exploratory in nature, the values between 0.05 and 0.1 are reported for descriptive purposes, as indicating trends for further investigation.

Results

Mean Familiarity Scores

Three types of means were computed for each predicted familiarity likelihood level ("likely," "somewhat likely," and "unlikely" to be familiar): total surface-level familiarity, GERD surface-level familiarity, and GERD concept-level familiarity (Table 2). Total surface-level familiarity reflects surface-level familiarity with terms on all three topics. Since the test included five terms per topic per level, 15 is the maximum possible surface-level familiarity score for each level. GERD surface-level familiarity indicates surface-level familiarity with GERD terms only, with five the maximum possible score (based on five GERD terms at each level). GERD concept-level familiarity reflects answers to GERD concept questions, with five the maximum possible score for each level.

Table 2. Mean surface-level and concept-level familiarity scores

Predicted Familiarity Likelihood	Total Surface-Level Familiarity Mean (SD)	GERD Surface-Level Familiarity Mean (SD)	GERD Concept-Level Familiarity Mean (SD)
Likely	13.80 (1.97)	4.75 (0.81)	3.83 (1.22)
Somewhat likely	12.92 (2.60)	4.54 (1.02)	3.94 (1.04)
Unlikely	9.53 (3.44)	3.42 (1.42)	3.04 (1.31)

Total surface-level familiarity and GERD concept-level familiarity were the dependent variables of hypotheses 1 and 2. GERD surface-level familiarity was used in computing the gap between GERD surface-level and concept-level familiarity, the dependent variable for hypothesis 3.

Predictors of Total Surface-Level Term Familiarity

Seven independent variables—predicted familiarity likelihood level, gender, English proficiency, highest education level, age, race, and health literacy level (S-TOFHLA scores)—were regressed onto the dependent variable, total surface-level term familiarity score. Linear regression found a statistically significant effect ($P < .001$) of predicted familiarity likelihood level on surface-level term familiarity. Health literacy was another statistically significant predictor of surface-level familiarity ($P < .001$). English proficiency was significant ($P = .05$); education level was not ($P = .15$).

Predictors of GERD Concept-Level Familiarity

All seven independent variables from the previous regression analysis plus GERD surface-level familiarity were regressed onto GERD concept-level familiarity score. Linear regression found statistically significant effects of predicted familiarity likelihood level ($P = .009$) and GERD surface-level familiarity score ($P < .001$) on GERD concept-level familiarity scores. The effect of health literacy level on GERD concept-level familiarity merits further investigation ($P = .06$).

XSL•FO
RenderX

Relating GERD Surface-Level and Concept-Level Familiarity Scores

While previous regression analysis indicated that GERD surface-level familiarity score was a significant predictor of GERD concept-level familiarity, the concept-level familiarity consistently lagged behind surface-level familiarity at all three levels (see Table 2). Linear regression analysis of the effect of predicted familiarity likelihood level on the surface-level–concept-level familiarity gap was performed. For the overall model, the gap was statistically significantly different from zero (P = .001). In addition, the gap was statistically significantly greater for terms predicted as "likely" then for those "not likely" to be familiar (P = .006). The gap for terms predicted as "somewhat likely" versus those predicted "not likely" to be familiar merits further investigation (P = .07).

Discussion

Implications for the Validity and Usefulness of the CHV Familiarity Model

Although preliminary in nature, this study presents an initial evaluation of the first model for estimating consumer familiarity with health-specific terms. The findings confirmed hypotheses 1 and 3 and partially confirmed hypothesis 2. Confirmation of hypothesis 1 provided initial validity evidence for the CHV familiarity likelihood model [8] by demonstrating a relationship between predicted familiarity and two types of empirically derived consumer familiarity scores. The brief "proof of concept" survey used in this study requires additional research to evaluate the underlying model's robustness with various target audiences of online consumer health materials: seniors, low-literacy individuals, chronic patients, etc. The approach used in the study provides a methodological framework for such follow-up validation studies. The present study, however, contributes to the field as it suggests that a health corpora frequency-based algorithm presents a feasible and more flexible alternative to general word lists or word length algorithms for estimating the difficulty of consumer health materials. For example, our existing model for predicting term difficulty can be used as a quick screening tool for determining "difficult" terms in consumer health texts and suggesting more consumer-friendly synonyms. Incorporating the model into a formula that produces a single text readability score would potentially automate the complex task of matching consumer health materials to readers (assuming that relevant reader information is available).

Insights for Improving the Power of CHV Familiarity Prediction

Partial confirmation of hypothesis 2 and confirmation of hypothesis 3 both point to limitations of the model with respect to its ability to identify "consumer-unfriendly" words. Part of the variance in readers' performance is likely to be related to demographic characteristics, not accounted for in the model. With further research, it is perhaps possible to adjust predicted familiarity likelihood categories for some target populations on the basis of known effects of demographics variables. However, identifying the full range of meaningful demographic variables is not realistic. Moreover, most sites are developed for a broad range of health consumers who represent a diverse range of competencies and experiences. This limitation is not unique to our approach but is true for all attempts to evaluate the difficulty of terms or a text. While individualized prediction of text difficulty on the basis of a model is desirable, it is also much more error prone than population-wide predictions because most predictive models are based on population statistics or empirical expert knowledge. Any prediction is necessarily an approximation, but a high-quality approximation is of considerable value. Presently, our predictive model framework also does not make a theoretical distinction between surface-level familiarity and conceptual understanding and does not make provision for the possible uneven gap between the two. If the uneven gap phenomenon is confirmed, then the "easiness" of terms predicted as highly likely to be familiar may be deceptive. Answering this question requires a strong operational definition of sufficient concept knowledge and a way of assessing it. The present instrument is an exploratory step in the direction of concept knowledge measurement. A satisfactory instrument should reconcile the goals of assessing a complex and multifaceted construct while being relatively quick and easy to administer.

Limitations of the Study

While most of the study results corresponded to our research hypotheses, the lack of significant effects of most demographic variables, particularly educational level, is surprising and may be due to sampling bias. It is possible that uneven representation obscured any education effects —41 out of 52 participants had at least some college education. Note that education is a proxy for general literacy, which is only one component of health literacy [10]. Other components, such as health care experience and motivation, may have a much stronger effect on health term familiarity and need to be explored in further research.

Follow-Up Work

Follow-up work includes validating and possibly adjusting the algorithm for specific populations, evaluating the role of potentially influential demographic variables in designs where these variables are represented across a broad range of values, and developing a formula that would assign a single-value text difficulty on the basis of the present algorithm. The calibration of such formulae in order to estimate the desired scores for various populations would require a set of extensive psychometric studies that are beyond the scope of most informatics research programs. However, developing the algorithm and testing its effectiveness against existing readability formulas are well within the capabilities of consumer health informatics research. It is also essential to develop methods to explore consumer understanding of health concepts in-depth, as the current study only touches the surface of this important topic.

XSL•FO
RenderX

Acknowledgments

This research was supported by the Intramural Research Program of the US National Library of Medicine, US National Institutes of Health (AK, TT, AB) and NIH grant R01 LM007222-05 (JC, LN, QZ). The authors thank Ilyse Rosenberg for her contribution to developing the instrument and Cara Hefner for assistance with the data collection.

Conflicts of Interest

None declared.

Multimedia Appendix

60-item questionnaire [PDF file (Adobe Acrobat), 64KB - jmir_v9i1e5_app1.pdf]

References

1. 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 May;285(20):2612-2621 [FREE Full text] [Medline: 11368735] [doi: 10.1001/jama.285.20.2612]
2. Eysenbach G, Powell J, Kuss O, Sa ER. Empirical studies assessing the quality of health information for consumers on the world wide web: a systematic review. JAMA 2002 May;287(20):2691-2700 [FREE Full text] [Medline: 12020305] [doi: 10.1001/jama.287.20.2691]
3. Chall JS, Dale E. Readability Revisited: The New Dale-Chall Readability Formula. Cambridge, MA: Brookline Books; 1995.
4. Plimpton S, Root J. Materials and strategies that work in low literacy health communication. Public Health Rep 1994;109(1):86-92 [FREE Full text] [Medline: 8303020]
5. Mccray AT. Promoting health literacy. J Am Med Inform Assoc 2005 Mar;12(2):152-163 [FREE Full text] [Medline: 15561782] [doi: 10.1197/jamia.M1687]
6. Chapman K, Abraham C, Jenkins V, Fallowfield L. Lay understanding of terms used in cancer consultations. Psychooncology 2003 Sep;12(6):557-566. [Medline: 12923796] [doi: 10.1002/pon.673]
7. Minsky M. A framework for representing knowledge. In: Winston PH, editor. The Psychology of Computer Vision. New York: McGraw-Hill; 1975:211-277.
8. Zeng QT, Tse T, Crowell J, Divita G, Roth L, Browne AC. Identifying consumer-friendly display (CFD) names for health concepts. AMIA Annu Symp Proc 2005:859-863. [Medline: 16779162]
9. Zeng Q, Kim E, Crowell J, Tse T. A text corpora-based estimation of the familiarity of health terminology. Proc ISBMDA 2005:184-192.
10. Baker DW, Williams MV, Parker RM, Gazmararian JA, Nurss J. Development of a brief test to measure functional health literacy. Patient Educ Couns 1999 Sep;38(1):33-42. [Medline: 14528569] [doi: 10.1016/S0738-3991(98)00116-5]
11. Lee SYD, Bender DE, Ruiz RE, Cho YI. Development of an easy-to-use Spanish Health Literacy test. Health Serv Res 2006 Aug;41(4 Pt 1):1392-1412. [Medline: 16899014] [doi: 10.1111/j.1475-6773.2006.00532.x]
12. Davis TC, Long SW, Jackson RH, Mayeaux EJ, George RB, Murphy PW, et al. Rapid estimate of adult literacy in medicine: a shortened screening instrument. Fam Med 1993 Jun;25(6):391-395. [Medline: 8349060]

Abbreviations

CHV: consumer health vocabulary
GERD: gastroesophageal reflux disease
SAHLSA: Short Assessment of Health Literacy for Spanish-Speaking Adults
S-TOFHLA: Short Test of Functional Health Literacy in Adults
REALM: Rapid Estimate of Adult Literacy in Medicine

XSL•FO
RenderX

submitted 30.11.06; peer-reviewed by L Slaughter; comments to author 08.12.06; revised version received 22.02.07; accepted 06.03.07; published 14.03.07.

Please cite as:
Keselman A, Tse T, Crowell J, Browne A, Ngo L, Zeng Q
Assessing Consumer Health Vocabulary Familiarity: An Exploratory Study
J Med Internet Res 2007;9(1):e5
URL: http://www.jmir.org/2007/1/e5/
doi:10.2196/jmir.9.1.e5
PMID:17478414

© Alla Keselman, Tony Tse, Jon Crowell, Allen Browne, Long Ngo, Qing Zeng. Originally published in the Journal of Medical Internet Research (http://www.jmir.org). Except where otherwise noted, articles published in the Journal of Medical Internet Research are distributed under the terms of the Creative Commons Attribution License (http://www.creativecommons.org/licenses/by/2.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited, including full bibliographic details and the URL (see "please cite as" above), and this statement is 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