The Mysterious Maze of the World Wide Web: What Makes Internet Health Information High Quality?

By
Joshua Seidman, PhD
Executive Director, Center for Information Therapy

Donald Steinwachs, PhD
Professor and Chair, Health Policy and Management Department, Johns Hopkins University, Bloomberg School of Public Health

Haya R. Rubin, MD, PhD
Professor of Medicine, Director, Quality of Care Research Departments of Medicine, Epidemiology, and Health Policy and Management, Johns Hopkins School of Medicine and Bloomberg School of Public Health

July 6, 2004
Contact us

To obtain information about these Center for Information Therapy resources:

- White paper: *The Ix Evidence Base: Using Information Therapy to Cross the Quality Chasm*, by Bianca R. Gwinn and Joshua Seidman, PhD
- White paper: *The Business Case for Information Therapy*, by Donald W. Kemper
- Book: *Information Therapy: Prescribed Information as a Reimbursable Medical Service*, by Donald W. Kemper and Molly Mettler
- To see examples of new information therapy innovations, read the 2004 Ix® Conference Report.
- Previous papers published in the peer-reviewed journal *The Journal of Medical Internet Research* provide considerable detail on the conceptual framework and the design and testing of the new tool described in this white paper:

Center for Information Therapy
A division of Healthwise, Incorporated
600 New Hampshire Avenue, NW
Washington, D.C. 20037
Phone: 202-945-6810
Fax: 202-266-6068
E-mail: jseidman@healthwise.org
www.informationtherapy.org
www.healthwise.org
Abstract

Consumers increasingly are turning to the Internet for help in making health care decisions, even though such information often is incomplete or inaccurate. Existing tools designed to help consumers find credible Internet health information focus almost exclusively on proxy measures of quality, such as the characteristics of the site sponsor and whether the site lists its sources. Our review of 90 sites with diabetes information found these proxy measures do little to explain the comprehensiveness and accuracy of the information provided. We lay out an alternative strategy for helping consumers find accurate and comprehensive Internet health information.

Introduction

Consumers are turning to the Internet in growing numbers for answers to their health care questions. As of 2002, 73 million Americans reported that they went to the Web for health information.1 At the same time, a number of studies have suggested that much of the health information available on the Internet is, to varying degrees, incomplete, inaccurate, oversimplified, and/or misleading.2 For example, a RAND study conducted in the fall of 2000 that assessed the quality of health care Web sites found that they were generally incomplete. Of 10 English-language sites for each condition, 36 percent had more than minimal coverage and complete accuracy for childhood asthma, 37 percent for obesity, 44 percent for depression, and 63 percent for breast cancer. Without an adequate medical background, consumers who rely on the Web for health information may seize on misleading, incorrect, or oversimplified information that can be potentially harmful to them.

Although some organizations have begun to develop tools that would help consumers navigate the health care information on the Web, there is not yet any widely accepted method for evaluating the quality of health and medical information on the Internet. Moreover, the existing tools generally do not attempt to look directly at the information that is provided on health care conditions and assess its accuracy and comprehensiveness. Instead, they evaluate the quality of health care Web sites through proxy measures of information quality that include how the Web site was developed and is operated (i.e., “structural” measures such as whether the Web site discloses the names of its authors, explains the process used to develop its content, and appropriately dates its content) and characteristics of the site’s sponsor (e.g., whether it is sponsored by a for-profit company). However, no research has ever addressed whether these proxy measures actually provide useful guidance to consumers about the accuracy and comprehensiveness of the information they are gathering from the Web.

No research has ever addressed whether these proxy measures [of information quality] actually provide useful guidance to consumers.

In an effort to investigate further the quality of health care information on the Web, as well as to evaluate the existing tools available to help consumers navigate health care information on the Web, we undertook a review of 90 diabetes Web sites using a new evaluation tool. As described in detail in the remainder of this paper, we found surprisingly little connection between the quality of information on diabetes Web sites and indicators consumers are advised to use to select reliable health information Web sites. In light of these findings, we conclude this article with a series of strategies and policy
recommendations for improving Web site accreditation efforts and providing more effective guidance to consumers.

We found surprisingly little connection between the quality of information on diabetes Web sites and indicators consumers often are told to look for.

Consumer Strategies for Navigating Health Information on the Web

In general, consumers show relatively little skepticism about the quality of health information on the Web. The Pew Internet & American Life Project found that 72 percent of the 73 million Americans who have gone online for health information say they can believe all or most of the health information online. At the same time, consumers will occasionally reject the information provided on the Web—close to three-quarters in the Pew study had done so at some point. When they do, it is for the following reasons:

- The site is "too commercial and seemed more concerned with selling products than providing accurate information" (47%).
- They couldn’t determine the source of the information (42%).
- They couldn’t determine when the site was last updated (37%).

The popular press reinforces the notion that sponsorship characteristics can help consumers find reliable Web sites. Despite a lack of evidence supporting this theory, advice articles in magazines such as Money and Ladies Home Journal generally tell lay people that they should look for sites that have no commercial interests, are run by government agencies, and/or are affiliated with respected academic institutions.

One of the country’s independent health care organization accrediting bodies, known as URAC, launched the first attempt at third-party Web site oversight in 2001. URAC’s voluntary, consumer Web site accreditation system relies on a set of ethical and quality standards that address several issues including privacy protection, security, and the process used for developing content. URAC’s effort represents an important step forward, but they also have substantial limitations in guiding consumers to credible health information. URAC’s quality standards are primarily designed to assess structural issues in Web site design and management and do not assess the specific quality or credibility of the information provided on the Web site.

Methods and Analysis Plan

For our research, we developed a new tool for evaluating the accuracy and comprehensiveness of information on the Internet. The tool (presented in Exhibit 2) includes a set of criteria that cover a cross-section of structural and performance measures. We took a number of steps to ensure the tool was systematic, objective, and based on measurable elements.

First, we aimed our tool at information on a specific condition, diabetes, for which the evidence base is stronger than in many other clinical areas due to specific, exceptional clinical trials (e.g., the Diabetes Control and Complications Trial [DCCT], the United Kingdom Prospective Diabetes Study [UKPDS], and the Diabetes Prevention Program [DPP]). As a result, it was possible for us to use evidence-based guidelines and performance measurement experts to develop a reasonable “gold standard” for the content and scope of information on diabetes that should be included in a Web site on the disease.

Second, we extracted elements from the American Diabetes Association’s largely evidence-based practice guidelines to identify the items that should be on a diabetes Web site (e.g., whether the site provides a comprehensive list of the potential complications of diabetes). We wrote a definition for each item in the tool in order to precisely specify what would constitute a positive score on each criterion. Third, we had three experts in diabetes performance measurement—all of whom served on the Diabetes Quality Improvement Project technical expert panel—review the tool. Fourth, we conducted a reviewer training session. Finally, the reviewers used an abstraction tool to evaluate the first 90 Web sites retrieved from a diabetes search using the Direct Hit
search engine. Considerably more detail on the tool development is described elsewhere.

Along with gathering information on the accuracy and comprehensiveness of health information on the 90 diabetes sites, we also gathered information for each site on the various proxy measures often used to evaluate Web site quality. For example, we gathered information on whether sites identified their authors and dated their content, as well as on the sponsorship characteristics of the site. By collecting these data, we were able to assess whether the structural measures actually provide consumers with accurate and comprehensive information on the quality of information on diabetes Web sites. We used Spearman rank correlations, which measure the strength of association between two variables, to do this.

We assessed the relationship between sponsorship characteristics and the accuracy and comprehensiveness of health information (using two-sample t-tests) specifically to answer the following questions: Which types of Web sites provide higher-quality health information:
- for-profit vs. not-for-profit
- government vs. private
- academic vs. nonacademic

Finally, controlling for the other factors (using multiple linear regression), we assessed which of these sponsorship characteristics actually has the greatest influence on the quality of health information Web sites.

The Quality of Diabetes Information on the Web

Enormous variation existed in the quality of diabetes information on the Internet (see Exhibit 1). The tool included 34 accuracy and comprehensiveness criteria; the quality was

Enormous variation existed in the quality of diabetes information on the Internet....Quality scores ranged from 14 to 97 percent among the sample of 90 Web sites evaluated.

Exhibit 1. Distribution of Diabetes Web Site Scores of Performance Measures
measured by the percentage of times a given site provided appropriate information according to the criteria. Quality scores ranged from 14 to 97 percent among the sample of 90 Web sites evaluated, with a median of 55 percent and a mean of 56 percent. The gap in quality was not just based on a few outliers; one-quarter of sites were below 41 percent, and one-quarter were above 70 percent.

In other words, on a quarter of the Web sites that have information on diabetes, a consumer would get inaccurate or incomplete information three out of every five times. For example, a consumer might have gone to one Web site sponsored by an academic institution that did not provide accurate information about the warning signs of an acute diabetic episode.

For a quarter of the Web sites that have information on diabetes, a consumer would get inaccurate or incomplete information three out of every five times.

Given the potentially dangerous consequences of such poor information and the reasonably high chance of getting it, there’s great benefit in determining the characteristics associated with high-quality Web sites. Such proxies for information quality could provide a valuable and simple strategy for steering consumers to credible information.

What Do Structural Measures Tell You About Web Site Quality?

In our analysis, we found that structural measures, similar to those developed by URAC, do offer some limited indication of diabetes Web site accuracy and comprehensiveness. There was a statistically significant but modest correlation between the performance measures (comprised of two individual components, comprehensiveness and accuracy) and structural measures. The Spearman rank correlation coefficient ($r_s$) was 0.42 (0.45 for comprehensiveness and 0.28 for accuracy).

While our findings suggest that there is some utility in considering structural measures of Web sites when assessing the quality of information that they provide, these measures alone are not sufficient to guide consumers to high-quality Web sites.

What Do Sponsorship Characteristics Tell You About Web Site Quality?

Contrary to the advice often offered by the popular press, our research found that sponsorship characteristics are a poor predictor of Web site quality in terms of the accuracy and comprehensiveness of health information. Statistical tests (two-sample t-tests) showed no statistically significant quality differences based on sponsorship. Not-for-profit sites (55 of the 90 sites), government-sponsored sites (12) and—somewhat surprisingly—non-academic sites (78) and sites that accept advertising (32) performed slightly better than those that did not accept advertising (58), although none of these differences were statistically significant.

Our research found that sponsorship characteristics are a poor predictor of Web site quality.

When controlling for other factors (using multiple linear regression), only one sponsorship characteristic was statistically significant in explaining the accuracy and comprehensiveness of health information: for-profit status. However, even that relationship was extraordinarily weak. As a whole, sponsorship characteristics could explain only 2 percent of the variation in accuracy and comprehensiveness after adjusting for expected chance prediction. (The $R^2$ was only 0.06, with an adjusted $R^2$ of only 0.02.) The average accuracy and comprehensiveness of for-profit sites was only 4 percentage points below that of not-for-profit sites, and scores of both ranged from below 20 percent to at least 90 percent. In other words, not-for-profit sites on average have higher-quality information than for-profit sites when holding other factors constant, but
the more important point is that the variation within each group far exceeds the predictive power of this independent variable.

Interestingly, one might have thought that for-profit status would be strongly correlated with other factors (particularly advertising and, to a lesser extent, government and academic sponsorship), and therefore, one might see the same impact of those characteristics. However, none of those independent variables statistically significantly affected Web site quality, whether or not one controlled for other factors.

What Makes Web Site Information High Quality?

The limited relationship between proxy measures of quality and the actual accuracy and comprehensiveness of diabetes information on the Internet suggests there is a compelling need to develop other methods for guiding consumers. This research demonstrates that one cannot simply use proxy measures or follow the advice common in the popular press to “just look for sites with certain characteristics.”

In our research, we also conducted a qualitative review of our findings to see if there was anything else that easily and readily explained variation in the quality of diabetes Web sites. Unfortunately, this review suggested that there are no ready alternatives to the proxy measures currently in use. We did find that three of the top four performing sites provided information exclusively on diabetes, but the highest-quality site and seven of the top ten provided information on a broader array of health conditions.

Our qualitative review suggested that some of the high-quality sites performed well because they simply took a more serious approach than others in developing information that would be helpful to consumers. For example, one site trains all of its medical writers in evidence-based medicine methodologies, uses trained medical librarians to conduct literature searches, has generalist physicians involved with the structuring of content, and has specialist physicians review all of the content when it is originally prepared as well as when it is updated. However, it would be difficult to objectively measure and verify whether a site uses these kinds of processes when developing its content.

Policy Implications

Our research indicates that the likelihood that consumers will find accurate and comprehensive diabetes content varies dramatically depending on what Web sites they view. It also suggests that it may be necessary to evaluate the content provided on each individual health Web site if we want to provide meaningful guidance to consumers about where they can obtain accurate and comprehensive information.

An Alternative Model

Although existing independent review bodies, such as URAC, provide useful information about some of the key proxy measures that might help lead people to higher-quality Web sites, our research suggests that structural measures do not correlate highly with performance in the realm of health information quality. Rather than completely replacing existing Web site accreditation efforts, the URAC structural measures should be supplemented by evaluations of information accuracy and comprehensiveness using tools such as the one we have developed for diabetes Web
sites. Performance measures aimed at evaluating condition-specific content would be complemented by information that addresses a Web site’s adherence to generic but important principles such as privacy concerns, ethics, and other structural characteristics.

If URAC moves in the direction of adding a review of information content, one of the major challenges will be to develop a practical and feasible strategy for extending the reach of a content-based tool that only offers information about one condition (diabetes) to a much broader array of health conditions. What mechanisms might exist to create comparable tools for a breadth of conditions?

The URAC structural measures should be supplemented by evaluations of information accuracy and comprehensiveness.

One could convene interested parties to train them in the process of developing a content-based tool and then have individual organizations dedicated to specific health conditions take responsibility for development and maintenance of a tool for their area. For example, the American Heart Association and American College of Cardiology might take responsibility for accreditation of Web sites with information about heart disease. Alternatively, the federal government might be able to play a key role; for example, the National Heart, Lung, and Blood Institute could oversee cardiovascular content. Given the enormous array of health conditions for which information is available on the Web, it would be important to establish priorities. Content-based accreditation efforts might begin with the conditions associated with the Centers for Disease Control and Prevention’s list of most common reasons for mortality or the Institute of Medicine’s recently published list of the 20 top priority health care areas for national quality improvement.

In many ways, our recommendation that it may be beneficial to add a review of actual content to Web site accreditation efforts is not surprising given the experience of other initiatives aimed at helping consumers assess the quality of health plans. For example, efforts to evaluate the quality of care provided by health plans undertaken by organizations such as the National Committee for Quality Assurance (NCQA) often began considering only structural measures. However, in 1999, NCQA began to base 25 percent of a health plan’s overall accreditation score on a set of clinical performance measures (known as HEDIS, the Health Plan Employer Data and Information Set) and member satisfaction measures (based on the consumer survey CAHPS, the national Consumer Assessment of Health Plans Study). Applying the same principles to Web site accreditation, performance measurement could begin primarily with a system similar to the diabetes tool presented here, but could ultimately incorporate a CAHPS corollary by adding a user satisfaction survey component that would assess user perceptions of Web sites’ usefulness, understandability, and other important issues.

Would it make a difference?

Given that such an approach will involve considerable effort, it is worth exploring the question: Will comparative Web site performance measurement make a difference in terms of improving the quality of Internet health information? If one looks to the quality-of-care measurement experience, there is some reason to think that report cards and publishing of comparative performance data do have an impact, although it may arise in a different way than might be expected. Whereas some had hoped that consumers and purchasers would use comparative performance data to select providers and health plans or that clinicians would use that data to make referrals, most evidence thus far suggests that has only occurred in limited ways.12

What comparative performance measurement and its public reporting have done, however, is to stimulate behavior change on the part of the organizations being profiled. This has happened at both the provider level, such as with the reporting of bypass surgery rates in New York state, and at the health plan level, with reporting of HEDIS data, as described in the NCQA State of Managed Care Quality Report.13 Given the corollaries in the development of performance measurement systems, public reporting of Web site information quality could have a similar impact.
Approaches to Quality Oversight

Creating a performance measurement-based system of consumer health Web site evaluation leaves many options available for how that information can be used. In a regulatory approach, Web sites would have to meet certain levels of accuracy and comprehensiveness in order to be allowed to deliver medical information to consumers. Some might argue that such limitations constitute an infringement on free speech in that the Internet is simply one vehicle for offering viewpoints and opinions that are constitutionally protected in the U.S. An opposing perspective could equate delivery of health information over the Internet to the delivery of that same information interpersonally during a clinical encounter. That is, a clinician requires state licensure to be able to deliver certain types of information to patients and can be sued for malpractice for delivering inaccurate health information to a patient. Some would argue that some Web sites purport to communicate comparable information without bearing the same responsibility.

In truth, such strict regulation of health Internet content seems unlikely and probably unconstitutional, but the findings from this research demonstrate that an objective mechanism of oversight is needed; several options for how that might be conducted are offered here. Such oversight could emanate from the private sector or some public-private partnership. Private and public sector entities may reimburse for information dissemination at some point in the future, in which case they would be able to mandate quality assurance for content delivered. Under such a system, either the government could fill that role itself or could deem existing or new private sector organizations appropriate for measuring and accrediting Web sites.

Future Research Directions

Three categories of future related research would be helpful. First, with respect to the need for research on Web site evaluation tools for other conditions, one of the critical factors is dealing with the varying evidence base across diseases. Whereas treatment for diabetes has a relatively strong evidence base (and treatments for some conditions like cardiovascular disease probably are even stronger in that respect), other conditions have much more limited or rapidly changing evidence on which a Web site can base its information. This has implications for criteria selection in terms of both what should be covered on a Web site (comprehensiveness) and precisely what the site should say (accuracy).

Second, more research on how consumers use consumer Web information would be valuable. Specifically, how do they search for information, what influences their use of the information that they find, and how does that information affect how they make health care decisions for themselves and their families? Although high-quality information is a prerequisite for effecting positive consumer health behavior, it is a necessary but not sufficient requirement for stimulating behavior change on the part of consumers.

Third, the research regarding how to integrate this tool with URAC’s existing accreditation effort probably could benefit from the experience NCQA has had with integrating performance measurement into an accreditation system based on structural measures. Specific issues to address include: how to weight the relative portions of each system in an overall score; what the timeline should be for system integration; what conditions should be addressed initially; how the specific conditions should evolve over time; and what parties should be involved in a multi-condition evaluation system.

Conclusion

Wide variation exists in the accuracy and comprehensiveness of online diabetes information, and there is no existing mechanism for consumers to get detailed, objective information about Web site quality. Furthermore, this research also demonstrates that proxies such as sponsorship characteristics and structural, descriptive issues are of limited use in helping consumers search for Internet health information. Objective review of performance in producing health information quality, expressed in terms of accuracy and comprehensiveness of information, can offer consumers a tangible and useful tool in navigating the online health universe.

Healthwise and the Healthwise logo are two of the registered trademarks of Healthwise, Incorporated. All other names and logos are (registered) trademarks of their respective holders. © Copyright 2004-2012, Healthwise, Incorporated. All rights reserved.

8119-100504

Page 7 of 9

www.informationtherapy.org


<table>
<thead>
<tr>
<th>Category</th>
<th>Measurement</th>
<th>How Measured</th>
</tr>
</thead>
<tbody>
<tr>
<td>I. Explanation of methods</td>
<td>a. Content generation explanation</td>
<td>a. Site has explanation of process for generating its health content</td>
</tr>
<tr>
<td></td>
<td>b. Identification &amp; disclosure</td>
<td>b. Author(s) listed and affiliations, credentials &amp; contact information provided</td>
</tr>
<tr>
<td>II. Validity of methods</td>
<td>a. Referenced material</td>
<td>a. Assertions supported by referenced material</td>
</tr>
<tr>
<td></td>
<td>b. Peer review</td>
<td>b. Material on site has gone through peer review</td>
</tr>
<tr>
<td>III. Currency of information</td>
<td>a. Updating process</td>
<td>a. Site has explanation for updating its health content</td>
</tr>
<tr>
<td></td>
<td>b. Content dating</td>
<td>b. Each Web page indicates date of last update</td>
</tr>
<tr>
<td></td>
<td>c. Timely content</td>
<td>c. Page updated within last 6 months</td>
</tr>
<tr>
<td>IV. Comprehensiveness of information</td>
<td>a. Screening</td>
<td>Each of these aspects (primarily drawn from the clinical practice recommendations of the American Diabetes Association) addressed and discussed on the Web site.</td>
</tr>
<tr>
<td></td>
<td>b. Glycemia tests</td>
<td></td>
</tr>
<tr>
<td></td>
<td>c. Nutrition</td>
<td></td>
</tr>
<tr>
<td></td>
<td>d. Exercise</td>
<td></td>
</tr>
<tr>
<td></td>
<td>e. Acute episodes</td>
<td></td>
</tr>
<tr>
<td></td>
<td>f. Secondary diabetes</td>
<td></td>
</tr>
<tr>
<td></td>
<td>g. Foot care</td>
<td></td>
</tr>
<tr>
<td></td>
<td>h. Dyslipidemia</td>
<td></td>
</tr>
<tr>
<td></td>
<td>i. Smoking cessation</td>
<td></td>
</tr>
<tr>
<td></td>
<td>j. Nephropathy</td>
<td></td>
</tr>
<tr>
<td></td>
<td>k. Retinopathy</td>
<td></td>
</tr>
<tr>
<td></td>
<td>l. Immunization</td>
<td></td>
</tr>
<tr>
<td></td>
<td>m. Insulin administration</td>
<td></td>
</tr>
<tr>
<td></td>
<td>n. Oral medications</td>
<td></td>
</tr>
<tr>
<td></td>
<td>o. Glucose monitoring</td>
<td></td>
</tr>
<tr>
<td></td>
<td>p. Care of children</td>
<td></td>
</tr>
<tr>
<td></td>
<td>q. Gestational diabetes</td>
<td></td>
</tr>
<tr>
<td></td>
<td>r. DCCT implications</td>
<td></td>
</tr>
<tr>
<td></td>
<td>s. UKPDS implications</td>
<td></td>
</tr>
<tr>
<td></td>
<td>t. Insulin/glucose explanation</td>
<td></td>
</tr>
<tr>
<td></td>
<td>u. Prevention</td>
<td></td>
</tr>
<tr>
<td></td>
<td>v. Psychological aspects</td>
<td></td>
</tr>
<tr>
<td></td>
<td>w. Neuropathy</td>
<td></td>
</tr>
<tr>
<td></td>
<td>x. Obesity</td>
<td></td>
</tr>
<tr>
<td>V. Accuracy of information</td>
<td>a. Type 1 vs. Type 2</td>
<td>1. Explain Type 1 (lack of insulin) and Type 2 (insulin doesn’t work effectively)</td>
</tr>
<tr>
<td></td>
<td>b. Secondary causes</td>
<td>2. Explain main secondary causes: drugs (pentamidine, corticosteroids, thiazides, niacin), pancreatic disease (chronic pancreatitis, hemochromatosis, cystic fibrosis, pancreatic surgery), endocrine disorders (Cushing’s disease, acromegaly, pheochromocytoma, thyrotoxicosis), genetic syndromes (lipoedystrophies, myotonic dystrophy, ataxia telangiectasia), insulin-receptor syndromes</td>
</tr>
<tr>
<td></td>
<td>c. Diagnostic tests</td>
<td>3. Explain diabetic threshold for fasting blood glucose test (&gt;125 mg/dl) and oral glucose tolerance test (&gt;199 mg/dl)</td>
</tr>
<tr>
<td></td>
<td>d. HbA1c test</td>
<td>4. Explain risk associated with HbA1c levels &gt;8%; impact on risk of CAD, kidney disease, and retinopathy.</td>
</tr>
<tr>
<td></td>
<td>e. Albumin tests</td>
<td>5. Explain macroalbuminuria test (goal: negative) and microalbuminuria test (goal: &lt;30 mg/g creatinine)</td>
</tr>
<tr>
<td></td>
<td>f. Cholesterol tests</td>
<td>6. Explain HDL/LDL difference and LDL target level (&lt;100 mg/dl)</td>
</tr>
<tr>
<td></td>
<td>g. Warning signs</td>
<td>7. Explain warning signs of acute diabetic episodes (fainting, seizures, state of serious confusion)</td>
</tr>
<tr>
<td></td>
<td>h. Hypoglycemia prevention</td>
<td>9. Explanation of all 5 classes of oral medications (sulfonylureas, meglitinitides, biguanides, glitazones, alpha glucosidase inhibitors)</td>
</tr>
<tr>
<td></td>
<td>i. Oral medications</td>
<td>10. Explain liver problems associated with the glitzone Rezulin and why pulled back from market</td>
</tr>
<tr>
<td></td>
<td>j. Rezulin</td>
<td></td>
</tr>
</tbody>
</table>
ENDNOTES


4 Ladies Home Journal. 2002 (April), 68.


