EDITORIAL

Infodemiology: The Epidemiology of (Mis)information

Gunther Eysenbach, MD

Much of the health information on the Internet has been described as being discordant with information from evidence-based sources (1). A new research discipline and methodology has emerged—the study of the determinants and distribution of health information and misinformation—which may be useful in guiding health professionals and patients to quality health information on the Internet. Information epidemiology, or infodemiology, identifies areas where there is a knowledge translation gap between best evidence (what some experts know) and practice (what most people do or believe), as well as markers for "high-quality" information.

The first infodemiological study was published in 1996 (2), but this type of research only became widely known with a subsequent publication in a prominent journal (3). A recent review identified 79 infodemiological studies (1), and as of today more than 100 articles have been published. Most of the early studies were descriptive, reporting the percentage of websites that had inaccurate or otherwise imperfect health information (1). Such studies are also useful in identifying where the evidence is conflicting, where fraud is prevalent, or where misleading advertisements prevail over balanced health education. For example, studies have concluded that up to 90% of information on diet and nutrition is unreliable, compared with only 5% for cancer (1).

Descriptive studies, however, do not explain how indicators for quality and website characteristics are related, and whether these characteristics are associated with utility to the consumer. Analytical studies, which employ statistical methods such as multivariate regression to explore how quality criteria and other variables are related (Figure), are more useful in addressing questions such as "Are government sites more accurate than commercial sites?" or "Are websites that disclose the authorship and include a date of last update more accurate?" One important question is which markers or characteristics of a website are "valid" quality criteria to discriminate or predict a "good" health website. A valid quality criterion would be a feature (or a combination of features) that predicts effective health communication in terms of improving knowledge or changing health behavior, or which is associated with a measurable effect on health outcomes. However, analyses of site characteristics and health outcomes can be complicated, since measuring the very small and indirect effect that a single site may have on health outcomes is not feasible. Furthermore, users rarely use only one website and often use other sources of information, and health outcomes may take years to develop.

Perhaps a more realistic aim of analytical studies is to answer the question of whether technical or formal site characteristics that suggest accuracy of content can be identified, as was done by Martin-Facklam et al. in this issue of the *Journal* (4). The authors found citing references and an absence of financial interest to be associated with content accuracy. Indeed, websites targeted at medical professionals are more likely to provide references (5) and may be more consistent with evidence-based guidelines. Disclosure of the authors' names was not found to be a predictor of site reliability, perhaps because government organizations (6), other organizations, and drug companies (7) do not usually provide the names of authors on their websites. Disclosure of when the website was last updated may also be an independent predictor of site accuracy, although this association was not observed in an earlier analysis (8). One study suggested that displaying the HONcode logo, having an organization (.org) domain, and displaying a copyright are predictors of content "reliability" (9). These authors, however, evaluated comprehensiveness rather than accuracy (1). For all these results, the question of whether these findings can be generalized arises.

Technical (or formal) markers are called "valid" quality criteria if they predict accurate content, as accurate content is assumed to lead to positive health outcomes. Conversely, failure to establish an association between a quality criterion and accuracy of site content does not mean that the marker is an "invalid" quality criterion, because some criteria such as disclosure of sponsorship or appropriate privacy policies are ethical tenets and quality criteria per se. They have face validity, independent of whether they predict an accurate website or a positive

Am J Med. 2002;113:763–765.
From the Centre for Global eHealth Innovation, University Health Network; and Department of Health Policy, Management and Evaluation, University of Toronto, Ontario, Canada.

Requests for reprints should be addressed to Gunther Eysenbach, MD, Centre for Global eHealth Innovation, Toronto General Hospital, R. Fraser Elliott Building, 4th Floor, Room 4S435, 190 Elizabeth Street, Toronto, Ontario M5G 2C4, Canada, or geysenba@uhnres.utoronto.ca.

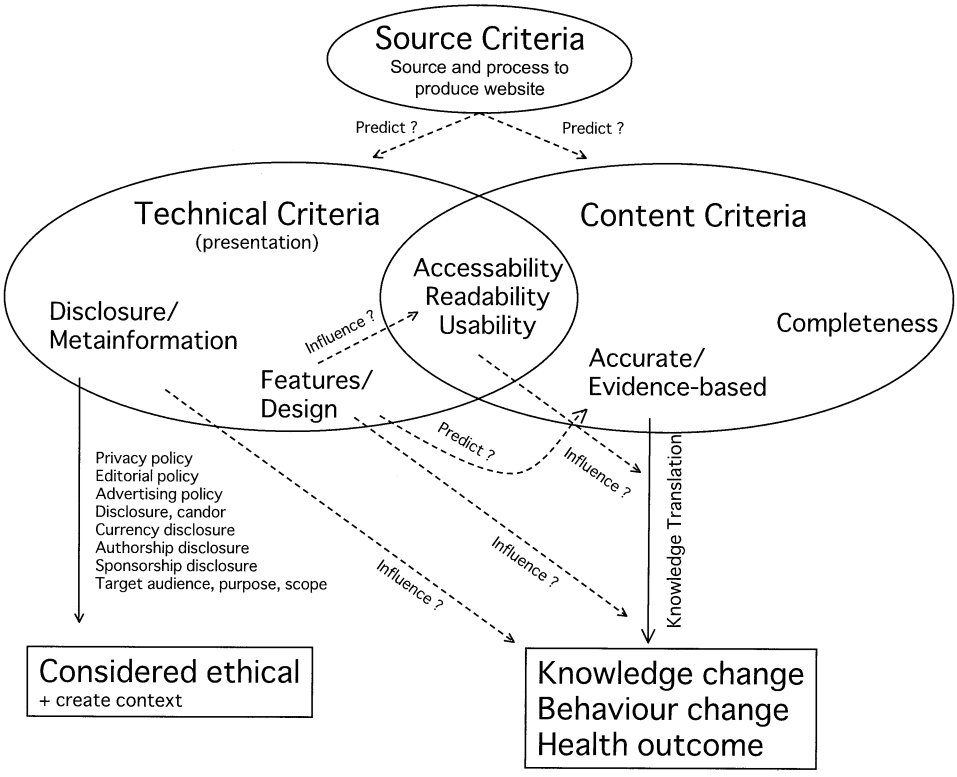

Figure. Conceptual framework of quality markers and their relation with outcome variables for effective health communication on the Internet.

health outcome, much as informed consent is a quality criterion for good clinical care. Other criteria have face validity because they create context; for example, disclosure of the target audience, scope, and purpose of the site helps users to filter inapplicable information; links and references enable users to verify the information given; and authorship credentials help users to decide if a site is trustworthy.

Analytical infodemiology studies can also be used to determine which design, content, and technical features influence the accessibility, readability and usability of a website, and how the knowledge translation process is affected subsequently. In addition, it is also useful to know whether and how source criteria (who produces the information and how) predict and affect the presentation and content of information. One study showed that websites owned by academic institutions were more accurate and better at providing accountability criteria (disclosure of author, copyright, date posted) (5). Another reported that organizations and sites with editorial boards were more complete and accurate than those owned by single persons (7). Government or nonprofit sites have been found to be more likely to display the date of creation than commercial sites (10), whereas websites for professionals are more likely to disclose when the information was posted, compared with sites for patients (11).

A number of initiatives have been developed to improve the quality of health information on the Internet (1). The U.S. Department of Health and Human Services aims to "increase the proportion of health-related World Wide Web sites that disclose information that can be used to assess the quality of the site" (12). Such information would include the identity of website developers and sponsors; how to contact the owners/developers of a site; potential conflicts of interest or biases; the explicit purpose of the site, including commercial purposes and advertising; original sources of content; how the confidentiality of personal information is protected; how the site is evaluated; and when content on the site is updated.

The MedCERTAIN project proposed a decentralized, open system based on promoting self-disclosure of information, similar to nutrition labeling for food, using a standardized format to describe, annotate, evaluate, and certify health information sites (13). This concept is currently being implemented in the MedCIRCLE project. Widespread use of the proposed metadata (information about information) vocabulary would allow more sophisticated analyses of the relation between website properties and quality, as well as the progress of the Healthy People 2010 objectives to be tracked.

At our institution, we have started a new initiative to

train patients to locate and assess health information on the Internet using the mnemonic "CREDIBLE":

- **C**urrent and frequently updated
- **R**eferences cited
- **E**xplicit purpose and intentions of the site
- **D**isclosure of developers and sponsors
- **I**nterests disclosed and not influencing objectivity (e.g., financial interests)
- **B**alanced content, lists advantages and disadvantages
- **L**abeled with metadata
- **E**vidence-level indicated

Further infodemiological studies will identify which quality markers should be added to this list to help users to select quality information, as well as developers to design websites that maximize the benefits of the Internet for improving public health.

REFERENCES

1. Eysenbach G, Powell J, Kuss O, Sa ER. Empirical studies assessing the quality of health information for consumers on the World Wide Web: a systematic review. *JAMA.* 2002;287:2691–2700.
2. Davison K. The quality of dietary information on the World Wide Web. *J Can Diet Assoc.* 1996;57:137–141.
3. Impicciatore P, Pandolfini C, Casella N, Bonati M. Reliability of health information for the public on the World Wide Web: systematic survey of advice on managing fever in children at home. *BMJ.* 1997;314:1875–1879.
4. Martin-Facklam M, Kostrzewa M, Schubert F, et al. Quality markers of drug information on the Internet: an evaluation of sites about St. John's wort. *Am J Med.* 2002;113:740–745.
5. Chen LE, Minkes RK, Langer JC. Pediatric surgery on the Internet: is the truth out there? *J Pediatr Surg.* 2000;35:1179–1182.
6. Berland GK, Elliott MN, Morales LS, et al. Health information on the Internet: accessibility, quality, and readability in English and Spanish. *JAMA.* 2001;285:2612–2621.
7. Griffiths KM, Christensen H. Quality of web based information on treatment of depression: cross sectional survey. *BMJ.* 2000;321:1511–1515.
8. Libertiny G, Perkins JM, Magee TR, Galland RB. Varicose veins on the Internet. *Eur J Vasc Endovasc Surg.* 2000;20:386–389.
9. Fallis D, Fricke M. Indicators of accuracy of consumer health information on the Internet: a study of indicators relating to information for managing fever in children in the home. *J Am Med Inform Assoc.* 2002;9:73–79.
10. Hoffman-Goetz L, Clarke JN. Quality of breast cancer sites on the World Wide Web. *Can J Public Health.* 2000;91:281–284.
11. Jiang YL. Quality evaluation of orthodontic information on the World Wide Web. *Am J Orthod Dentofacial Orthop.* 2000;118:4–9.
12. *Healthy People 2010: Understanding and Improving Health.* 2nd ed. Washington, D.C.: Office of Disease Prevention and Health Promotion, U.S. Department of Health and Human Services; 2000. Publication 017-001-00547-9.
13. Eysenbach G, Yihune G, Lampe K, et al. MedCERTAIN: quality management, certification and rating of health information on the Net. *Proc AMIA Symp.* 2000:230–234.