

Descriptions of Web Sites in Search Listings: A Potential Obstacle to Informed Choice of Health Information

Michael D. Slater, PhD, MPA, and Donald E. Zimmerman, PhD

The American public increasingly uses the Internet to obtain health information.¹⁻⁵ Researchers have scrutinized the quality of

health Web sites,⁶⁻¹⁰ but relatively few have studied the Web portals that deliver sites to users.^{11,12} Health consumers must rely on brief site listings generated by portal searches when deciding which Web site to access. These analyses examine the extent to which such site descriptions provide adequate cues about site sponsors and their intent to sell products on Web sites.

METHODS

A content analysis was conducted on Web sites generated in response to 24 search terms selected in consultation with health education experts concerning heart disease, cancer, and weight loss. Heart disease terms included *heart disease, heart attacks, heart attack prevention, heart attack treatment, heart disease cure, blocked arteries, angina, and cardiac arrest*. Cancer terms included *cancer, cancer prevention, cancer treatment, cancer cure, skin cancer, breast cancer, prostate cancer, and lung cancer*. Weight loss terms included *weight loss, diets, dieting, weight control, cellulite, losing fat, burn fat, and overweight*.

Each of the 24 terms was searched on the 5 most widely used portals at the time of the study in late fall 2001 (Yahoo!, Microsoft Network [MSN], America Online, Lycos, and Go),¹³ and the first 10 sites listed were coded, for a total of 1200 Web sites analyzed (minus 13 eliminated because of data entry errors). Coding was restricted to the first 10 listings because prior research indicates that lay users rarely go beyond the first screen or two of listings.^{12,14} This sample, therefore, represents the sites likely to be encountered in a lay search on these health topics at the time of study.

The first 12.5% of the sites were double-coded to establish reliability. Another 12.5% were double-coded midway through the analysis to test coder drift. Cohen κ values for intercoder reliability were 0.76 (first iteration) and 0.73 (second iteration) for whether or not text accompanying the site listing indicated that a Web site was selling products or services. Descriptive information included text with the site listing, text available when the cursor "rolled over" the site listing, and the uniform resource locator (URL) itself. It proved more difficult to reliably ascertain

whether the organization sponsoring the Web site was adequately identified in the site listing description or URL, with a kappa of 0.52 for the first coding iteration and 0.63 for the second. This lower reliability reflected ambiguities in the site descriptions. Kappa is a highly conservative statistic with reliabilities of 0.5 deemed adequate for analysis and report.^{15,16}

RESULTS

Overall, 30.2% of the site listings indicated that the corresponding Web site was selling a product. Analyses reported elsewhere determined that 39.1% of these Web sites were actually selling products.¹⁷ Therefore, 23% of the Web sites selling products (8.9% of all sites coded) were not identifiable as such based on the site listings. The difference between the number of sites selling products and those that were identifiable as such from the site listings was highly significant ($t_{1137}=8.17$, $P<.001$).

Search listings normally began with one or more paid commercial listings, typically identified as “sponsored” sites. Most portals included a link to a disclaimer defining “sponsored” sites as paid advertisers. Only America Online included the disclaimer in the listing itself. MSN listed “featured” sites, including both recommended sites and advertisers.

Moreover, only 46.9% of the listings identified the organizational sponsor of the Web site, although the precision of this estimate must be qualified given the difficulty in unambiguously assessing whether the source could be adequately identified from the listing description.

There were no significant differences in the number of listings that clearly identified the source or in discrepancy between commercial sites located and those clearly listed, by search portal or by health topic. These were tested by using analysis of variance with search term as the unit of analysis, nested within health topic, and with search portal as a within-subjects factor.

DISCUSSION

Listings of search results provided by the most widely used Web portals often do not

provide basic information a consumer would need to select an objective and reliable health information Web site. Search descriptions of health sites ideally should permit consumers, before actually accessing a Web site, to determine more consistently and accurately the source of the information and whether the site is selling products or providing information free of commercial intent. Commercial sites are of particular concern given evidence that the large majority of such sites promote unregulated supplements and unproven remedies and services¹⁷ and that pseudoscientific claims made by such promotional sites may be persuasive even to readers with college-level scientific training.¹⁸ Future research should examine how diverse members of the public use information on site source, commercial intent, and commercial sponsorship of the listing when selecting Web sites that they will use to obtain health information. ■

About the Authors

The authors are with Department of Journalism and Technical Communication, Colorado State University, Fort Collins. Michael D. Slater has a joint appointment in the Department of Psychology.

Requests for reprints should be sent to Michael D. Slater, PhD, MPA, Department of Journalism and Technical Communication, Colorado State University, Fort Collins, CO 80523-1785 (e-mail: michael.slater@colostate.edu).

This brief was accepted December 28, 2002.

Contributors

M.D. Slater was primarily responsible for overall study design, data analysis, and drafting of the brief. D.E. Zimmerman collaborated in study design and writing of the brief, oversaw data collection, and made technical determinations regarding choice of search engines and search procedures. Responsibility for developing the coding scheme and training coders was shared equally.

Acknowledgments

This research was supported in part by the Center for Research in Writing and Communication Technology, Colorado State University.

The authors thank Christel Irvin and Stephanie Sheely for their assistance with coding.

Human Participant Protection

No protocol approval was needed for this study, as no human participants were involved.

References

1. Cline RJW, Haynes KM. Consumer health information seeking on the Internet: the state of the art. *Health Educ Res*. 2001;16:671–692.

2. Fox S. Wired seniors [Pew Internet & American Life Project Web site]. September 9, 2001. Available at: <http://www.pewinternet.org/reports/toc.asp?report=40>. Accessed November 14, 2001.

3. Landro L. More people are using Internet health sites, but fewer are satisfied. *Wall Street Journal*. December 29, 2000:9.

4. Spooner T, Rainie L. Hispanics and the Internet [Pew Internet & American Life Project Web site]. July 25, 2001. Available at: <http://www.pewinternet.org/reports/toc.asp?report=38>. Accessed November 14, 2001.

5. Spooner T, Rainie L. African-Americans and the Internet [Pew Online Life Report Web site]. October 22, 2001. Available at: <http://www.pewinternet.org/reports/toc.asp?report=25>. Accessed November 14, 2001.

6. Baur C, Deering MJ. Proposed frameworks to improve the quality of health Web sites [review]. *Online Medscape Gen Med* [serial online]. 2000;2(3):1–5. Available at: <http://www.medscape.com/viewarticle/418842>. Accessed April 21, 2003.

7. Beredjikian P, Bozentka D, Steinberg D, Bernstein J. Evaluating the source and content of orthopaedic information on the Internet: the case of carpal tunnel syndrome. *J Bone Joint Surg Am*. 2000;82-A:1540–1543.

8. Gordon J, Barot L, Fahey A, Matthews M. The Internet as a source of information on breast augmentation. *Plast Reconstr Surg*. 2001;107:171–176.

9. Griffiths K, Christensen H. Quality of Web based information on treatment of depression: cross sectional survey. *BMJ*. 2000;321:1511–1515.

10. Latthe M, Latthe P, Charlton R. Quality of information on emergency contraception on the Internet. *Br J Fam Plann*. 2000;26:29–43.

11. Berland GK, Elliot MC, Morales LS, et al. Quality of health information on the Internet: accessibility, quality, and readability in English and Spanish. *JAMA*. 2001;285:2612–2621.

12. Eysenbach G, Köhler C. How do consumers search for and appraise health information on the World Wide Web? Qualitative study using focus groups, usability tests, and in-depth interviews. *BMJ*. 2002;324:573–577.

13. Sullivan D. Jupiter media matrix search engine ratings. October 5, 2001. Available at: <http://www.searchenginewatch.com/reports/mediamatrix.html>. Accessed October 29, 2001.

14. Nielsen J. Search: visible and simple. May 13, 2001. Available at: <http://www.useit.com/alertbox/20010513.html>. Accessed August 18, 2002.

15. Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics*. 1977;33:159–174.

16. Perreault WD, Leigh LE. Reliability of nominal data based on qualitative judgments. *J Marketing Res*. 1989;26:35–48.

17. Slater MD, Zimmerman DE. Characteristics of Web sites identified by major search portals. *JAMA*. 2002;288:316–317.

18. Haard J, Slater MD, Long M. Scientese and ambiguous citations in the selling of unproven medical treatments. *Health Commun*. In press.