

Evaluating the prevalence, content and readability of complementary and alternative medicine (CAM) web pages on the Internet

Smitha Sagaram M.B.B.S., Muhammad Walji B.S., Elmer Bernstam, M.D., M.S.E.

School of Health Information Sciences, The University of Texas Health Science Center at Houston, Houston, TX

Abstract

Complementary and alternative medicine (CAM) use is growing rapidly. As CAM is relatively unregulated, it is important to evaluate the type and availability of CAM information. The goal of this study is to determine the prevalence, content and readability of on-line CAM information based on searches for arthritis, diabetes and fibromyalgia using four common search engines. Fifty-eight of 599 web pages retrieved by a "condition search" (9.6%) were CAM-oriented. Of 216 CAM pages found by the "condition" and "condition + herbs" searches, 78% were authored by commercial organizations, whose purpose involved commerce 69% of the time and 52.3% had no references. Although 98% of the CAM information was intended for consumers, the mean readability was at grade level 11. We conclude that consumers searching the web for health information are likely to encounter consumer-oriented CAM advertising, which is difficult to read and is not supported by the conventional literature.

Introduction

Complementary and alternative medicine (CAM) includes treatments and health care practices not widely taught in medical schools and not generally used in hospitals¹. The National Center of Complementary and Alternative Medicine extends this definition to include practices that are not usually reimbursed by medical insurance companies².

In 1997, it was estimated that four in 10 Americans visited a CAM practitioner spending \$27 billion (not reimbursed)³. The major forms of CAM include acupuncture, herbs, homeopathy, therapeutic massage, and traditional oriental medicine.²

In a study assessing patients' perception of CAM therapies, 79% of those who saw a doctor and used CAM, "perceived the combination be superior to either one alone."¹

Many studies have looked at patient-oriented health information available on the Internet, often focusing

on specific health conditions. Berland et al. (2001) concluded that the coverage of health information is poor and inconsistent although accuracy is generally good. They also noted that a high reading level is required to comprehend the information.⁴ Li et al. (2001) reviewed web sites related to back pain and found that most were classified as advertising and information quality varied considerably. Of the 74 web sites, only 9 were considered high quality.⁵ The authors did not differentiate between CAM and conventional web sites. On their review of carpal tunnel syndrome web sites, Beredjiklian et al (2000) found the information to be of "limited quality and poor informational value". Fewer than half the sites were reported to provide "conventional information," while 23% presented "unconventional information."⁶ Hellowell et al. (2000) reported that most urology-related web sites provided conventional and good quality information. The authors noted "that there were reassuringly few 'unconventional' web sites".⁷ Few studies specifically address information content or readability of CAM web sites.

Many studies that have peripherally included CAM web pages have branded them "unconventional" and implied that they are inaccurate. However, it is important to assess the prevalence and evaluate the content of CAM web pages as patients visiting them may be more likely to take direct action from the information they receive. Furthermore, consumers are unlikely to disclose their use of CAM therapies to their physicians.¹ Unlike conventional medicine, consumers have unregulated access to many CAM therapies like herbs and vitamins. Therefore, the quality of CAM web pages may be even more important as they may directly influence consumer behavior.

The goal of this study was to evaluate the likelihood that a consumer looking for health information on the web will encounter CAM information. We examined search results returned from queries on three common health conditions using 4 popular search engines. We also evaluated the content and readability of CAM

web pages in order to find out who produces them, for what purpose, and how easy they are to read.

Methods

The health conditions 'diabetes', 'arthritis' and 'fibromyalgia' were chosen as the condition search keywords as they are among those conditions most frequently searched by consumers on MEDLINE-*plus*.⁸ Diabetes is a condition that is successfully managed by conventional medicine, while arthritis and fibromyalgia are less well managed by conventional medicine.

Four of the more popular search engines; Google, Hotbot, AltaVista and Excite were chosen as the media to carry out the searches. These search engines were chosen as they are among the most visited⁹ and have also been used in other studies that evaluated health information on the Internet.^{4,5,10}

In order to determine prevalence of CAM web sites, the search phrases: 'diabetes,' 'arthritis' and 'fibromyalgia' were each searched in all four of the search engines. The first 50 results from each search engine results page were analyzed. In order to gather more CAM web pages to determine the content and readability of the CAM web pages, an additional search term: "herbs" was appended to the condition search phrase (condition + herbs). These additional queries were also searched on all four search engines, and the first 20 results from each were used.

Two reviewers were arbitrarily assigned to two different search engines, and used the same exact queries to procure the search results.

A web based data entry page linked to a Microsoft Access database was created to facilitate capturing and categorizing the search results.

Methods to find Prevalence

Classification of web pages

The first 50 results of web pages in each search engine were classified into one of the following categories:

1. CAM: The web page predominantly listed information or products associated with CAM treatments and therapies for humans.
2. Conventional: The web page predominantly listed information or products associated with conventional treatments for humans

3. Integrative: The web page discussed both CAM and conventional treatments and therapies for humans

4. Foreign: The web page was not in English

5. Other: The web page discusses a non-health related aspect, or is veterinary related

6. Unrelated: The web page is unrelated to the intent of the search phrase

Methods for Content Analysis of CAM Web Pages

Sites that were identified as CAM-oriented in the prevalence search were subjected to further analysis. An additional search query (condition + herbs) was also run for all three conditions in all 4 search engines to identify more CAM web pages (using the first 20 results from each search engine) as per the above classification. Each CAM web page was analyzed for the following:

Authorship of the web page

Each search result was classified as being authored by one of the following (adapted from Soot et al.1999¹⁰)

1. Academic: Those individuals or departments associated with a university, private educational research institution or not for profit related organization.
2. Government: Those whose web page extension culminated in .gov
3. Commercial: Those web pages that were marketing products or services.
4. News Authors: Those web pages that carried news stories, identifiable by media logos or web page intent
5. Personal Authors: Individual authors who could not be identified as physicians or CAM practitioners and who did not represent an academic, government, commercial or news institution.
6. Physician Authors: Individual physician authors who did not represent an academic, government, commercial or news organization.
7. CAM Practitioners: Individuals who identified themselves as CAM practitioners who did not represent an academic, government, commercial or news organization.
8. Unidentifiable: Those web pages whose authorship could not be identified.

Audience

Each web page's main audience was identified by classifying them into one of the following:

1. Consumer: The web pages content included information or product offerings for consumers or those who were suffering from the particular condition
2. Professional: The web pages content was directed at a professional; physician, CAM practitioner, educator or health care professional.
3. Unclear: The web pages audience was unidentifiable, or it was not obvious whether the main audience was a consumer or a professional

Purpose

Each web page's main purpose was identified as one of the following:

1. Consumer Information: Health related information aimed at a consumer that was not linked with a product or service
2. Professional Information: Information aimed at a health care professional that was not linked with a product or service
3. Commercial to Consumer: Product or service based information aimed at a consumer that was not associated with health related information
4. Commercial to Professional: Product or service based information aimed at a professional that was not associated with health related information
5. Consumer Information and Commercial: Health related information aimed at a consumer that was linked with a product or service
6. Professional Information and Commercial: Health related information aimed at a professional that was linked with a product or service
7. For Research: The main purpose of the web page is related to funding of research.
8. Unclear: The main purpose of the web page is unclear
9. Unrelated: The main purpose of the web page is unrelated to intent of the search query.

Type of References

Each web page was analyzed to determine the type of reference source the information was based on. The types of reference sources were (adapted from Soot et al. 1999¹⁰):

1. Conventional: Cited literature that could be examined and validated like books or journal articles.
2. Anecdotal: Those web pages that predominantly made claims and references based on the authors experiences or beliefs
3. No References: Those web pages that had lacked references that had information that was customarily cited
4. Cannot be referenced: Those web pages that would not be customarily referenced in scientific publications (such as product listings)

Kitemarks

Each web page was analyzed for the appearance of a health related kitemark (e.g. HON Seal¹¹), often used to demonstrate the trustworthiness of the information presented.

Readability Analysis

Those CAM web pages that contained substantial (a paragraph or more of prose) CAM related information were further analyzed for readability using the Flesch-Kincaid (FK) reading level. Reviewers were permitted to go up to two links deep to find relevant material. The main content of each web page was copied and pasted into Microsoft Word 2002 and the computerized FK Grade level was recorded. The FK readability formula is calculated as follows: $(0.39 \times \text{average number of words per sentence}) + (11.8 \times \text{average number of syllables per words}) - 15.59 = \text{US grade level of education required to read a passage.}^{12}$ The FK formula reflects the number of words per sentence and the number of polysyllabic words in a sentence, but does not consider technical language or jargon.¹²

Results: Prevalence of CAM web pages

Five hundred and ninety-nine web pages were found for all three condition searches in the 4 search engines. There were 9.6% (58) CAM related web pages, 69.6% (417) conventional, 6.7% (40) Integrative, and the rest were found to be foreign, other or unrelated. The prevalence of CAM web pages by disease were 11.1%, 4.5%, and 13.5% for arthritis, diabetes, and fibromyalgia respectively. Among the 58 total CAM pages retrieved, 46% came from the 'fibromyalgia' search, 38% from 'arthritis' and only 16% from the 'diabetes' search. The search engine Excite returned 43% of the 58 total CAM web pages found, while AltaVista returned 26%, Hotbot 21% and Google 10%.

Results: Content of CAM Web Pages

The following results were obtained for CAM web pages found through both the 'condition specific' and 'condition + herbs' searches.

Authorship

Most of the CAM web pages had a commercial author (78.2%). Academic and personal authors made up 6%, and CAM practitioners and unidentifiable authors were 4.2%. Physician and government authors each had less than 1% representation.

Similarly, condition-specific searches yielded 87%, 78% and 71% commercial authors for arthritis, diabetes and fibromyalgia respectively.

Audience

Of the CAM related web pages reviewed 98% were aimed at consumers. Similarly, 99%, 94%, 99% of the web pages for arthritis, diabetes and fibromyalgia respectively were consumer based. Per search engine, 98%, 98%, 95% and 98% of the results returned by AltaVista, Excite, Google and Hotbot respectively were consumer oriented.

Purpose

Overall, 69.4% of the CAM web pages reviewed contained a commercial aspect. Commercial to consumer and commercial to professional web pages accounted for 44% of the results. Of all CAM web pages reviewed, those that provided solely consumer information constituted 29%, while 25% of the web pages contained both consumer and commercial information. There were few sites dedicated in providing professional information (1.9%), commercial to professional (0.5%) or both professional and commercial information (0.5%)

Fibromyalgia and arthritis commercial web pages made up 51.1% and 40.3% respectively, while 36.7% of the diabetes web pages were commercial.

References

Conventional references were cited in 24% of the CAM web pages, while 52.3% had no references at all. Anecdotal references made up 9.7%, and 13.9% were unable to be referenced. Searches for 'arthritis', 'diabetes' and 'fibromyalgia' specifically revealed that 58.4%, 40.8% and 53.3% respectively of the web pages were not referenced.

Kitemarks

Of the 216 CAM web pages reviewed only 8 (3.7%) were found to have Kitemarks displayed on the web

page, despite a study documenting 47 rating instruments in 1998.¹³

Readability

Of the 216 CAM web pages, 158 (73%) were analyzed for readability level as they contained information that could be assessed. The mean readability level was found to be at Flesch-Kincaid reading grade 11. Fibromyalgia web pages had a mean readability of 10.7, and arthritis and diabetes web pages were at 11 and 11.6 respectively.

Discussion

We found that a user is likely to encounter a CAM web page about 10% of the time when searching for arthritis, diabetes or fibromyalgia. Under our search conditions, a user searching for 'fibromyalgia' or 'arthritis' would encounter a CAM web page more often than a search for 'diabetes'. We also found that search engines differed in the proportion of CAM web pages returned in response to the same query. Of the 58 CAM web pages identified by the condition specific searches, Excite returned 43%, while Google returned only about 10%.

Of the 216 CAM web pages identified through the condition-specific search and the appended "condition + herb" search, 78% were found to have commercial authors. The purpose of these web pages was also found to be either commercial (44%) or both consumer and commercial information (25%). This indicates that under our search conditions, a consumer is likely to come across a CAM web page produced by commercial organizations whose intent is to sell a CAM related product or service. Interestingly, government based sites that may be considered more trustworthy, were encountered only 0.5% of the time.

The vast majority (98%) of CAM web pages were intended for consumers. However the readability analysis showed a mean reading grade level of 11. This is far higher than grade 8 which is recommended for consumer health information.¹⁴ This seems to be similar to what others have found while analyzing conventional medicine web sites.⁴

A quarter of the CAM web pages contained conventional references, while over half did not contain any references despite having information that would normally be cited. Contrastingly, a study on Vascular Surgery on the Internet found that most authors (60.9%) referenced conventional sources of information.¹⁰ This may be explained by the fact that many CAM therapies are considered unproven. Very few (3.7%) of the CAM web pages were found to have kitemarks, such as the HON Seal, that are supposed

to be emblems of trustworthiness. The credibility and usefulness of kitemarks has been brought into question¹³, which may explain its low usage.

Under our particular search conditions, the results showed a difference in the number of CAM pages depending on the health condition searched, but further studies would need to be done to ascertain whether this difference is repeatable and statistically significant. An overwhelming majority of the CAM pages were produced by a commercial entity whose purpose was also commercial. Further studies can look at the accuracy and quality of the content appearing on CAM web pages.

Limitations of our study include the absence of inter-observer reliability. Our search results may not appropriately represent all CAM related web pages as we looked at a small subset of results, and no single search engine has been able to index all the possible pages. However, evidence suggests that consumers are likely to only look at the first set of search results.¹⁵ Further research is needed to test the generalizability of our results to other conditions as we sampled only three common conditions.

Consumers searching the web for health information are likely to encounter consumer-oriented CAM web pages, commercial in nature, which are difficult to read and are not referenced in a conventional manner. Inaccurate CAM information on the Internet may pose a greater health risk than conventional information and should be investigated further rather than being ignored.

References

1. Eisenberg DM, Kessler RC, Van Rompay MI, Kaptchuk TJ, Wilkey SA, Appel S, Davis RB. Perceptions about complementary therapies relative to conventional therapies among adults who use both: results from a national survey. *Ann Intern Med* 2001; 135(5): 344-351.
2. <http://nccam.nih.gov/an/general/> Accessed February 6, 2002. NCCAM General Information
3. Eisenberg DE, Davis RB, Ettner SL, Scott A, Wilkey S, Van Rompay M, Kessler RC. Trends in alternative medicine use in the United States, 1990-1997. *JAMA*. 1998; 280: 1569-1575.
4. Berland GK, Elliott MN, Morales LS, Algazy JI, Kravitz RL, Broder MS, Kanouse DE, Munoz JA, Puyol JA, Lara M, Watkins KE, Yang H, McGlynn EA. Health information on the Internet: accessibility, quality, and readability in English and Spanish. *JAMA* 2001; 285(20), 2612-2621.
5. Li L, Irvin E, Guzman J, Bombardier C. Surfing for back pain patients: the nature and quality of back pain information on the Internet. *Spine* 2001; 26(5), 545-557.
6. Beredjikian PK, Bozentka DJ, Steinberg DR, 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 (11), 1540-1543.
7. Hellawell GO, Turner KJ, Le Monnier KJ, Brewster SF. Urology and the Internet: an evaluation of Internet use by urology patients and of information available on urological topics. *BJU Int* 2000; 86(3), 191-194.
8. Lacroix E-M. Health Topics most Hit March 2000, 2000.
9. <http://searchenginewatch.com/reports/netratings.html>, Nielsen/NetRatings Search Engine Ratings
10. Soot LC, Moneta GL, Edwards JM. Vascular surgery and the Internet: a poor source of patient-oriented information. *J Vasc. Surg* 1999; 30(1), 84-91.
11. <http://www.hon.ch/HONcode/Conduct.html> Accessed February 6, 2002. Hon Code of Conduct (HON code) for medical and health Web sites: Principles: Health on the Net Foundation.
12. Oermann MH, Wilson FL. Quality of care information for consumers on the Internet. *J Nurs Care Qual* 2000; 14(4), 45-54.
13. Jadad AR, Gagliardi A. Rating Health Information on the Internet: navigating to knowledge or to Babel. *JAMA* 1998; 279(8): 611-4
14. Streiff L, "Can Clients Understand Instruction?" *Image: Journal of Nursing Scholarship* 18 1986: 48-52
15. Eysenbach G, Kohler 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; 9; 324(7337): 573-7.