Quality of Health Information for Consumers on the Web: A Systematic Review of Indicators, Criteria, Tools, and Evaluation Results

Yan Zhang
School of Information, University of Texas at Austin, 1616 Guadalupe Street, Austin, TX 78701.
E-mail: yanz@ischool.utexas.edu

Yalin Sun
School of Information, University of Texas at Austin, 1616 Guadalupe Street, Austin, TX 78701.
E-mail: clairesun05@utexas.edu

Bo Xie
School of Nursing and School of Information, University of Texas at Austin, 1710 Red River Street, Austin, TX 78701. E-mail: boxie@utexas.edu

The quality of online health information for consumers has been a critical issue that concerns all stakeholders in healthcare. To gain an understanding of how quality is evaluated, this systematic review examined 165 articles in which researchers evaluated the quality of consumer-oriented health information on the web against pre-defined criteria. It was found that studies typically evaluated quality in relation to the substance and formality of content, as well as to the design of technological platforms. Attention to design, particularly interactivity, privacy, and social and cultural appropriateness is on the rise, which suggests the permeation of a user-centered perspective into the evaluation of health information systems, and a growing recognition of the need to study these systems from a social-technical perspective. Researchers used many preexisting instruments to facilitate evaluation of the formality of content; however, only a few were used in multiple studies, and their validity was questioned. The quality of content (i.e., accuracy and completeness) was always evaluated using proprietary instruments constructed based on medical guidelines or textbooks. The evaluation results revealed that the quality of health information varied across medical domains and across websites, and that the overall quality remained problematic. Future research is needed to examine the quality of user-generated content and to explore opportunities offered by emerging new media that can facilitate the consumer evaluation of health information.

Introduction

The web is now the single largest source of health information for consumers. However, unlike traditional media, it is unregulated territory. Anyone with basic web programming skills can quickly launch a serviceable website. Thus, while embracing the enhanced accessibility to health information, almost all parties involved in healthcare have expressed concerns about the quality of health information online, despite few cases of harm having been documented and reported in the literature (Crocco, Villasis-Keever, & Jadad, 2002). Nevertheless, due to the potential detrimental consequences that poor-quality health information could cause, attention to the quality issue has not subsided. In past decades, numerous reviews synthesizing the literature on this subject have been published. Among them, Jadad and Gagliardi (1998) and Gagliardi and Jadad (2002) focused on identifying quality-rating instruments that were used to produce awards or seals of approval for websites; Bernstam, Shelton, Walji, and Meric-Bernstam (2005) focused on instruments that could be used by consumers; whereas Kim, Eng, Deering, and Maxfield (1999) surveyed criteria proposed to evaluate health-related websites. A commonality of these reviews is that they included academic publications, but mainly emphasized instruments or criteria that could be easily accessed through the open web. As a result, they shed limited light on how quality was defined and evaluated in evidence-based research publications.

Eysenbach, Powell, Kuss, and Sa (2002) filled this gap by reviewing 79 studies containing quantitative data on structure and process measures of the quality of health information on

the web. Their review touched on aspects such as evaluation criteria, the methodological rigor of studies, and evaluation results, with limited attention to the instruments used in the studies. Since its publication there has been an explosion in the amount of health information and an increase in the number of people affected by such information. As of 2012, 72% of web users looked online for health information, up from 25% in 2000 (Pew Research Center, 2013). Moreover, the past decade has seen the emergence of a wide array of new technologies, particularly Web 2.0 social, which have transformed the ways in which people seek health information. Instead of passively receiving information, people have begun to share experiences, participate in conversations with similar others, and collaboratively construct medical and health knowledge (Chou, Prestin, Lyons, & Wen, 2013; Eysenbach, 2008; Fox, 2011; Hamm et al., 2013). In this new social-technical environment, the definitions of quality and evaluation criteria may have evolved, and the quality of health information may has undergone changes.

This review examines how quality is currently defined and evaluated. Specifically, we set out to identify the criteria and instruments that researchers are using to evaluate the quality of consumer-oriented health information on the web and to synthesize the evaluation results. To achieve an objective or evidence-based view of the current landscape of health information quality, we limited the sources to peer-reviewed journal articles that document detailed evaluation processes. The knowledge gained from the review can also be expected to shed light on potential means to assist consumers in evaluating health information.

Literature Review

The quality of health information concerns every stakeholder involved in healthcare, particularly since the web has made access to this information unprecedentedly easy. These concerns are reflected in a continual stream of empirical studies, reviews, syntheses, debates, and commentaries (e.g., Cline & Haynes, 2001; Eysenbach et al., 2002; Lam, Roter, & Cohen, 2013; Post & Mainous, 2010; Silberg, Lundberg, & Musacchio, 1997; Wyatt, 1997). A major focus of these articles was to examine and synthesize the criteria used for evaluation. For example, Silberg et al. (1997) identified four sets of objective criteria for judging the quality of website content by referring to criteria applicable to the print world: authorship (authors and contributors, their affiliations, and relevant credentials), attribution (references and sources for all content, and copyright), disclosure (ownership, sponsorship, underwriting, commercial funding arrangements, and potential conflicts of interests), and currency (the date on which content was posted and updated). By reviewing published rating tools and journal articles, Kim et al. (1999) extracted 165 criteria and grouped them under 13 categories: content, design and aesthetics, disclosure, currency, authority, ease of use, accessibility and availability, links, attribution, intended audience, contact address or feedback mechanism, user support, and miscellaneous. Some of these criteria, such

as ease of use, intended audience, and user support, signify recognition of the subjective nature of quality.

In a systematic review, Eysenbach et al. (2002) grouped quality criteria into five categories: technical, design, readability, accuracy, and completeness/comprehensiveness; the technical category encompassed criteria from Silberg et al. (1997). Using the lens of persuasive computing (Fogg, 2003; Tseng & Fogg, 1999), O'Grady (2006) further grouped these technical criteria into four categories: presumed credibility, surface credibility, earned credibility, and reputed credibility. While the existence of the large number of criteria and different ways to categorize them is a reflection of the complexity of the concept, it also indicates that there remains a lack of consensus on what constitutes quality and what the major dimensions are.

This lack of conceptual clarity contributes to a major practical challenge: the lack of effective tools to facilitate quality evaluation. In practice, various combinations of quality criteria have been packed into instruments or tools to guide the evaluation of health information. In 1997, Jadad and Gagliardi (1998) identified 47 instruments that had been used to produce awards, quality ratings, or seals of approval, or to declare a site as meeting quality standards or receiving top ratings. Fifty-one new instruments were identified (Gagliardi & Jadad, 2002). In 2005, the number increased to 273 (Bernstam et al., 2005). Nevertheless, few of these instruments were for practical for use by consumers, as most either failed to disclose criteria, contained far too many elements, consisted of elements that could not be evaluated objectively (thus leading to low reliability), or were too complex and time-consuming to apply (Bernstam, Sagaram, Walji, Johnson, & Meric-Bernstam, 2005). Approaches have been proposed to reduce the consumers' burden, including providing third-party validation and quality seals (e.g., the HON code) (O'Grady, 2006; Risk & Dzenowagis, 2001), and creating software that automatically detects quality indicators on websites (Eysenbach & Diepgen, 1998; Wang & Liu, 2007). However, limited success has been achieved. Evaluating health information remains one of the biggest challenges facing ordinary consumers (Arora et al., 2008; Car, Lang, Colledge, Ung, & Majeed, 2011; Feufel & Stahl, 2012). At the same time, the purpose, value, and validity of quality-rating instruments have also been frequently questioned (Barnes et al., 2009; Hendrick et al., 2012; Khazaal et al., 2010; Wilson, 2002).

To achieve a better understanding of the concept of quality and to shed light on the development of tools facilitating evaluation, it is necessary to examine how quality is perceived and evaluated in existing studies. In this paper we review published, peer-reviewed empirical studies that have used predefined criteria to evaluate the quality of consumer-oriented health information on the web.

To provide a conceptual framework to the review, we adopted a decision-making perspective, viewing the evaluation and judgment of the quality of health information as decision-making activity. Decision-making theories (Brunswik, 1952; Stanovich, 1999) posit that judgment and

decision making are cognitive processes that involve information acquisition and processing. In information acquisition, individuals collect relevant information cues from the environment. In information processing, they interpret, compare, contrast, and evaluate the acquired information against certain criteria and make decisions based on the evaluation (Brunswik, 1952; Hogarth, 1987). In this process, collected information cues serve as indicators of criteria and reflect the value embedded in the criteria (Brunswik, 1952). Following these notions, it can be argued that when evaluating online health information, people collect information cues, or indicators, from the information environment, and evaluate the indicators against certain criteria, that reflect values important to the information seeker. Therefore, in this review, unlike most prior reviews, we distinguish quality indicators and criteria, hoping to bring a greater level of clarity to the concept of quality. Consistent with our practical goal to inform the development of tools and strategies that can facilitate users in evaluating information, we also review tools employed in the selected studies and synthesize the evaluation results.

Methods

Inclusion and Criteria

Table 1 lists the inclusion criteria for filtering articles for the review. One purpose is to provide an evidence-based view of the quality of consumer-oriented health information on the web; therefore, studies included were empirical, from peer-reviewed journal articles, evaluated information oriented toward consumers (as opposed to professionals), and provided detailed descriptions of the evaluation criteria and process. Reviews, commentaries, or articles published in journals but lacked sufficient details about the evaluation process (e.g., abstracts and research letters), were excluded. To be included, the evaluators of the information needed to be health professionals, subject experts or information professionals, and the information was evaluated against certain criteria (e.g., authority and accuracy) or certain instruments (e.g., DISCERN and HONcode). Articles that reported consumers' evaluations of online health information were excluded, because such results represent consumers' perceptions of quality and may be biased, compared to quality measured by clear, technical standards.

Because of the sheer number of articles that met the criteria, we also excluded papers not in English. The publication dates of the article were limited to the time period between January, 2002 and the time that the search was performed (January, 2013), because the previous systematic review on this subject, by Eysenbach et al. (2002), was published in 2002.

Search strategy

Two major search strategies were used. First, we searched 13 major academic databases. Second, we forward-tracked papers that cited Eysenbach et al. (2002), the most recent systematic review of the quality of online health information for consumers.

Database search. Table 2 lists the databases searched. They were chosen because they are major databases that cover health-related subjects. The search query was: (quality OR credibility OR reliability OR accuracy OR readability OR evaluation OR assessment) AND (health information) AND (online OR Internet OR web OR ehealth OR e-health OR cyber* OR electronic) AND (criteria OR criterion). Based on the size of the results returned, we modified the search by four different combinations of search field limitations in different databases. Specifically, in PubMed the query was searched by full text; in IEEE Xplore, ACM Digital Library, ScienceDirect, and Springer Link the query was searched by the search field "title"; in the Web of

TABLE 1. Inclusion criteria.

Aspects	Inclusion criteria
Type of publication	Peer-reviewed journal articles that report on empirical studies and document a detailed evaluation process
Information evaluated in studies	Information accessible on the Web
Target audience of the evaluated information	Consumers
Evaluators	Health professionals, subject experts, or information professionals
Methods of evaluation	Authors used certain instruments (e.g., DISCERN) or criteria (e.g., comprehensiveness and currency) to evaluate the information.
Language of the evaluated information	English
Publication dates	January 2002–January 2013

TABLE 2. Database search.

	Database	No. of records retrieved
1	ACM (Association for Computing Machinery) Digital Library	623
2	ArticleFirst	2
3	EBSCOhost Research Databases	805
4	Emerald Insight	202
5	IEEE Xplore	1,000*
6	IngentaConnect	124
7	JSTOR	56
8	ProQuest Databases	430
9	PubMed	3,192
10	ScienceDirect	1,000*
11	Springer LINK	1,000*
12	Web of Science	5,719
13	Wiley Online Library	125

Note. *Due to the citation exporting limitation set by these databases, only the first 1,000 records on the results lists, which were sorted by relevance, were downloaded.
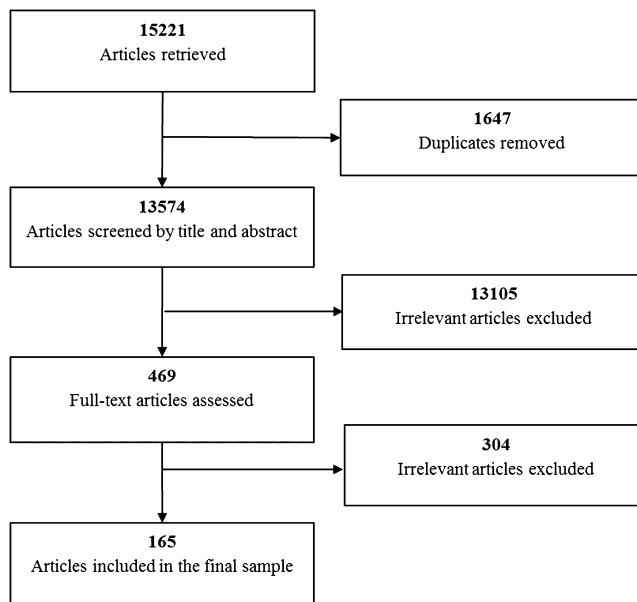
```
              ┌─────────────────────┐
              │       15221         │
              │  Articles retrieved │
              └─────────────────────┘
                        │
                        │        ┌─────────────────────────┐
                        ├───────▶│          1647           │
                        │        │    Duplicates removed   │
                        │        └─────────────────────────┘
              ┌─────────────────────────────────┐
              │              13574              │
              │ Articles screened by title and  │
              │            abstract             │
              └─────────────────────────────────┘
                        │
                        │        ┌────────────────────────────┐
                        ├───────▶│           13105            │
                        │        │ Irrelevant articles excluded│
                        │        └────────────────────────────┘
              ┌─────────────────────────────────┐
              │               469               │
              │  Full-text articles assessed    │
              └─────────────────────────────────┘
                        │
                        │        ┌────────────────────────────┐
                        ├───────▶│            304             │
                        │        │ Irrelevant articles excluded│
                        │        └────────────────────────────┘
              ┌─────────────────────────────────┐
              │               165               │
              │ Articles included in the final  │
              │            sample               │
              └─────────────────────────────────┘
```

FIG. 1. Process of identifying articles for review.

TABLE 3. Data extraction.

Category	Item
Basic article information	— Title — Year of publication — Author — Author affiliation — Article's country of origin (based on authors' affiliation) — Authors' academic domain/discipline — Year the study was conducted — Study location
Methods for selecting websites for review	— Credentials/position of the evaluators — Search terms — Search tools used — Web page selection strategies — Sample size — Medical domains evaluated — Objects being evaluated (web pages or websites)
Methods for evaluation	— Evaluation criteria — Instruments used — Quality indicators — Whether the article categorized web pages, if so, what categories were used — Rating scale — Number of raters — Inter-rater reliability
Study results	— Key findings — Methodological limitations — Authors' conclusions

Science, the query was searched against two search fields: "topic" or "title"; and in the remaining databases the search was by three search fields: "title," "abstract," or "keywords."

Forward chaining. Using a forward chaining strategy, we tracked papers that cited the seminar paper Eysenbach et al. (2002), using Google Scholar. In all, 941 articles were retrieved and combined with those retrieved from the database search.

A total of 15,221 records were identified from these two searches. Guided by the inclusion criteria, two independent reviewers each screened the title and/or abstract of a subset of these articles. Obviously irrelevant articles, such as those regarding the development of medical devices and those assessing medical technologies, therapies, and healthcare services, were excluded. This preliminary examination resulted in 469 articles for further review. The two reviewers then each scanned a subset of the full text of these articles and excluded those that did not meet the inclusion criteria. The final sample included 165 articles. The filtering process is portrayed in Figure 1.

Data Extraction

A data extraction form was created, consisting of 25 items grouped into four categories, shown in Table 3. For each article, a Word file (Microsoft Office 2007) was created to store the extracted information. One reviewer read the full text of all the selected articles and extracted the necessary information. To check the reliability of the extraction, a second reviewer extracted information from a randomly selected subset of 48 articles. The interrater agreement rate was 94.2%. Disagreements were resolved by discussion.

Data Analysis

The Word documents generated from the data extraction were imported to Nvivo 10.0 (QSR International), a content analysis software application, for further analysis. Nvivo was chosen because it can easily group documents with the same value on a certain data point (e.g., articles that used Google as a search tool). Moreover, it supports both quantitative and qualitative data analysis, which gave us the flexibility to code data at a more fine-grained level when needed. A reviewer read each document and coded the text into appropriate categories and subcategories. The top-level categories were the items in the data extraction form (Table 3). Under each category, subcategories were created when needed. For example, under the category selecting websites for review selection—search tools used, three subcategories were created: search engines, medical portals, and meta-search engines.

Most extracted data were straightforward and could be assigned to categories and subcategories without much ambiguity. The exceptions were criteria, indicators, and the overall tone of an article's conclusions. To guide the coding, "criteria" and "indicators" were defined as follows (Zhang, 2014):

• *Criteria:* abstract rules by which the quality of information is judged. Criteria reflect the values held by the evaluator regarding what is important for determining the quality of health information.
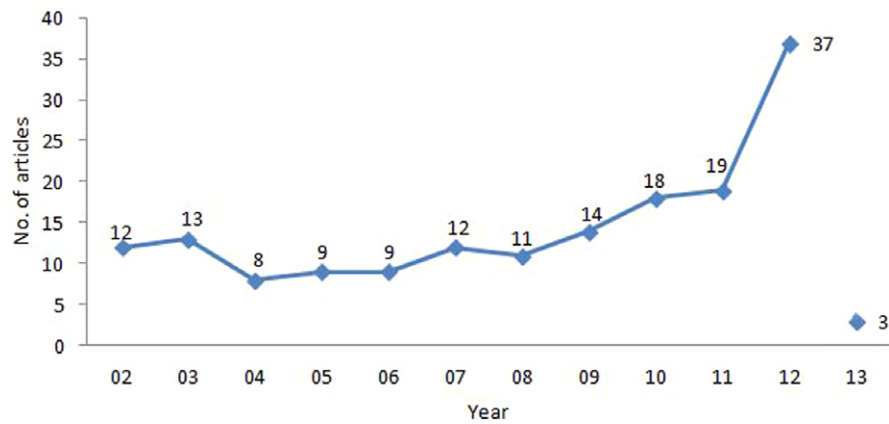
FIG. 2. The distribution of articles in the final sample (January, 2002 to January, 2013). [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

- *Indicators:* observable attributes associated with a website or its content, which serve as clues to whether or not a website or content meets a given criterion.

Rules for coding the overall tone were adapted from Eysenbach et al. (2002). An article was coded as

- *Negative,* if authors were being pessimistic or wary about about the quality of the information they evaluated;
- *Positive,* if the authors spoke positively, or did not express concerns;
- *Neutral,* if the authors discussed both risks and benefits and concluded with neither a positive nor negative tone;
- *Varied,* if the authors explicitly concluded that the information evaluated varied in quality.

Two reviewers each coded a subset of the documents, with 20% overlap (33 randomly selected articles). The intercoder reliability, calculated by the percentage of agreement, reached 90.2%. Disagreements were resolved by discussion.

Results

Basic Characteristics of the Included Articles

The final sample included 165 articles published between January, 2002 and January, 2013. The number of publications shows a generally increasing trend over time, with the peak in 2012, as shown in Figure 2. Because only 1 month of data was available in 2013, the line linking to 2013 was omitted to avoid a biased view.

These articles originated from 11 countries (based on authors' affiliations), with the top four being the U.S. (which contributed 37.6% of the articles, $n = 62$), UK (22.4%, $n = 37$), Canada (10.9%, $n = 18$), and Australia (8.5%, $n = 14$). Approximately 127 distinct medical subject areas were covered, including specific medical conditions (e.g., diabetes, cancer, and epilepsy), tests, medications, treatments (e.g., liver transplantation and pediatric surgery), nutritional information (e.g., malnutrition and

TABLE 4. Characteristics of the research methods of the included articles.

	Number of articles	Percentage
Search methods		
Search tools used		
Search Engines	158	95.8%
Meta-search engines	12	7.3%
Medical portals or specific websites	22	13.3%
Number of search engines used		
Only one	40	24.2%
More than one	122	73.9%
Keywords		
Generated by the authors	159	96.4%
Generated by consumers	6	3.6%
No. of web pages/sites included (Mean = 78.5; SD = 73.4)		
<50	71	43.0%
50~100	50	30.3%
>100	39	23.6%
Unknown	5	3.0%
Evaluation		
No. of independent raters		
One	25	15.2%
More than one	99	60.0%
Unknown	41	24.8%
Inter-rater reliability		
Not mentioned	105	63.6%
Only offered an overall reliability score	44	26.7%
Offered the inter-rater reliability score for every cruteruib in an instrument	16	9.7%

Mediterranean diet), and alternative medicine. The methods adopted by the researchers are shown in Table 4.

Most articles relied on general web search engines, mainly Google, Bing, and Yahoo!, to locate web pages for evaluation, with the majority using multiple engines. A dozen articles used meta-search engines (e.g., Dogpile and Metacrawler). The remainder evaluated information from

medical information portals (e.g., WebMD, MedlinePlus, Healthfinder, PatientUK, and NHSDirect Online). The search terms were mainly generated by researchers to simulate terms that consumers would use. Only six articles used terms generated by consumers. These were drawn from questions posted by consumers in online forums, prior published literature, focus groups, or patient surveys.

Most articles clearly defined the inclusion and exclusion criteria when selecting websites for evaluation. Generally, websites were included when they were (a) in the top (8–200 items) of the search results list, (b) relevant to the topic of interest, (c) intended for general consumers, and (d) written in English. Sources were excluded when they were topically irrelevant, duplicates, or not intended for general consumers. Some authors imposed more strict exclusion criteria. For example, some excluded user-generated content (e.g., blogs, Wikipedia, and discussion forums), sources sponsored by Google or by pharmaceutical companies, or sources in the form of news articles or videos. It is worth noting that, in most articles, authors did not explicitly differentiate web pages and websites, with the majority evaluating only web pages listed in search engines' results lists and a small portion extending the evaluation to relevant web pages under the same domain names. The number of web pages/sites evaluated in the studies ranged from three to 388 (M = 78.5; SD = 73.4) (except for two outlines, one of which evaluated 1,800 and the other 2,400 web pages). In their evaluations, 60.0% of the articles involved multiple evaluators. Of these, 60.1% reported interrater reliability.

Quality Criteria and Indicators

Table 5 summarizes the quality evaluation criteria used in the selected articles, and their corresponding indicators. Eleven criteria were identified. Based on the elements being evaluated, they can be grouped into two major categories: (a) criteria evaluating the content, and (b) criteria evaluating the design of technological platforms hosting the content. Content-related criteria can be further classified into those concerning the substance of the content and those concerning its formality. Two criteria, accuracy and completeness, were associated with the substance. The indicators for these were mostly derived from domain-specific medical guidelines, textbooks, or literature. For example, in evaluating the accuracy and completeness of information for common foot and ankle diagnoses, Meric et al. (2002) examined five areas of information (disease summary, pathogenesis, diagnostic tests, treatment options and complications, and prognosis and outcomes) against two American Academy of Orthopaedic Surgeons (AAOS) sanctioned textbooks (The *AAOS Comprehensive Orthopaedic Review* and *Orthopaedic Knowledge Update 4: Foot and Ankle*). Tsunehara and Gaster (2002) evaluated the accuracy of information concerning diabetes by comparing the content being evaluated with guidelines published by the American Diabetes Association.

Three criteria were associated with the formality of content: currency, credibility, and readability.

- *Currency* refers to whether or not the content is up-to-date. The main indicators include the publication date and the time of the last update. Twenty-five articles simply mentioned "up to date" as an indicator without specifying what it refers to.
- *Credibility* consistent with its definition in persuasion research (O'Keefe, 2002), has two components, authoritativeness and trustworthiness. The former refers to whether or not the content was contributed by people or institutes with authority; the latter refers to whether a source is truthful or biased. These two components were closely related and were often gauged by the same indicators (e.g., a highly regarded author). Most indicators for credibility identified in this review overlapped with the technical criteria identified in Eysenbach et al. (2002) and the objective criteria identified in Silberg et al. (1997), and they were grouped into three categories: attribution, authorship, and disclosure. Three additional categories of indicators were identified: the presence of third-party accreditations (mainly HON certificate), site popularity, and the presentation of the content.
- *Readability* refers to whether or not the content of a site is understandable for general consumers without medical background. In most cases, it was measured by one, or a combination of, the following tests: Flesch Reading Ease, Flesch-Kincaid the Grade Level, and the Simple Measure of Gobbledygook (SMOG). These tests help estimate the years of education needed to understand a piece of writing. The calculation was often carried out by a computer program, mostly the "Readability Statistics" tool in Microsoft Word or an online readability calculator.

Design-related criteria included accessibility, a esthetics, navigability, interactivity, privacy and data protection, and social and culture appropriateness.

- *Accessibility* refers to whether or not a site can be easily accessed. Widely used indicators included whether: a site is available, the links are active, special software is required for viewing the content, website contact information is clearly presented, and the site attends to users with disabilities.
- *Aesthetics* refers to the look and feel of a site. The major indicators are site layout (e.g., whether the layout is easy-to-follow, attractive, clear, simple, clean, and appealing), the use of images (e.g., whether they are relevant, appropriate, useful, and of high quality), and the use of headings (whether headings and subheadings are used).
- *Navigability* refers to how easily a consumer can move around within a site. Major indicators include whether the information architecture of a site is logical, supports easy navigation, and provides a site map.
- *Interactivity* refers to the capacity of a site to allow users to communicate with the system or with other users. Commonly used indicators include whether the site offers internal search functions, supports user input (e.g., commenting on content, and e-mails) and information exchange (e.g., chat rooms, bulletin boards, forums, and links to social media), provides multimedia content, and personalizes

TABLE 5.   Quality criteria and indicators.

Elements being evaluated		Criteria	Indicators (Number of articles)	
Content	Substance	Accuracy	Varied across domains (123)	
		Completeness	Varied across domains (121)	
	Formality	Currency	— Publication date (108)	— Copyright date (2)
			— Time of last update (84)	— Date of next update (1)
			— Sites respond to current events (7)	
		Credibility (authoritativeness & trustworthiness)	Authorship — Author name and professional credentials (124) — Editorial process (95) — Site domain and site type (12) Disclosure — Aims of the website (97) — Owner or sponsor of the site (80) — Financial disclosure and interest conflict (79) — Contact information disclosed (77) — Advertising policy (72) — Target audience disclosed (68) — Bias disclosed (11) Attribution — Source of the content and references (135) — Additional source of support (48) — Copyright, logo or page title disclosed (42) — Links to other related sites (16)	Third party accreditation — HON-Certified (37) Site popularity — Page rank in search engine results list (6) — Site traffic statistics (2) — High number of inlinks (1) Presentation of the content — Balanced content (51) — Spelling errors (8)
	Design	Readability	Facilitated by automatic tools (62)	
		Accessibility	— Operational: sites available, no dead links, and browser independent (59) — Clear presentation of website contact information (29) — Registration and accessing fee (16)	— Accessible for people with disabilities (e.g, font size and graphics with captions) (10) — Other languages offered (6) — Website technical support available (5) — Time to load a page (4) — Text-only option (2)
		Aesthetics	— Site layout (28) — Appropriate images use (26) — Use of headings (16) — Color schema (12)	— Design consistency (10) — Distracting graphic or text displays (5) — Fonts (4)
		Navigability	— Navigation structure: information presented in a logical order & easy navigation between links (40) — Site map (15)	— Easy to return to home page or go back (10) — Always tells your current location (7) — Consistency of menu systems (4)
		Interactivity	— Information exchange (e.g., forums, e-mails) (45) — Internal search engines (39) — Multimedia capability (21)	— FAQ section (14) — Personalization (14) — Social media expansion (3)
		Privacy & data protection	— Policy on the collection and use of personal data (40)	— Can use site anonymously (1)
		Cultural sensitivity	— The logic, language, and experience of the content match those of the intended audience (3)	— Presents cultural images and examples in realistic and positive ways (3)

content based on consumer characteristics. It is worth noting that, consistent with the rapid diffusion of social media, several studies measured whether a site actively pushed content to consumers by using social media sites like Facebook and Twitter (e.g., Mousiolis, Michala, & Antsaklis, 2012).

- *Privacy and data protection* refers to whether a site respects the privacy and confidentiality of personal data submitted by visitors. Most studies used the indicator outlined in HON-code's privacy criterion, that is, the presence of policy statements describing what information is collected and how it is used—for example, whether users were given the opportunity to opt out of sharing personal information (e.g., e-mail and personal health information). One study measured privacy in terms of whether users can use the site in an anonymous manner.

- *Cultural sensitivity* refers to whether the design of a site is culturally appropriate for its users. Indicators include surface elements (e.g., formats, pictures, and language) and latent messages and themes (e.g., the provision of examples to patients from diverse sociodemographic backgrounds). Among the selected articles, only three measured cultural sensitivity. Two of them used the cultural appropriateness subscale from the Suitability Assessment of Materials (SAM), and one used the Cultural Sensitivity Assessment Tool (CSAT) to conduct the assessment.

TABLE 6. Quality evaluation instruments.

	No. of instruments	Number of articles	Percentage
General evaluation instruments			
Preexisting instruments	29	112	67.9%
Compiled instruments	63	63	38.2%
Domain-specific evaluation instruments			
Based on medical guidelines	63	63	38.2%
Authors' expertise	N/A	43	26.1%

Quality Evaluation Instruments

Types of instruments used in evaluation. The instruments used to conduct quality evaluation can be classified as general or domain-specific. The former were used across different subject domains, and the latter were applied only to a particular domain (e.g., diabetes and eating disorders). Table 6 shows the usage of the two types of instruments in the selected articles.

Because the general instruments were not domain-specific, it is not surprising that they were mostly used to evaluate factors associated with the formality of content and/or the design of websites. These instruments were sometimes preexisting and sometimes compiled specifically for a study. Among the selected articles, 112 (67.9%) used a total of 29 different preexisting instruments, such as DISCERN, HONcode, and JAMA benchmarks. The compiled instruments were used in 63 articles (38.2%). Their construction was based on discussion among authors, or accomplished by selecting and combining criteria/indicators from prior studies and/or preexisting instruments. About a dozen studies used both preexisting and compiled instruments.

The domain-specific instruments were used almost exclusively to evaluate the substance of content (i.e., its accuracy and completeness). In 63 articles (38.2%), the instruments were created on the basis of medical guidelines, textbooks, or literature, whereas in 43 articles (26.1%), the authors, many of whom were domain experts, or invited domain experts, evaluated the content based on their medical expertise.

Most-used preexisting instruments among the selected articles. Of the 29 preexisting instruments, 15 (51.7%) were used in only 1 article, 3 (10.3%) were used in 2 articles, and 2 (6.9%) were used in 3 articles. The remaining eight (27.6%) were used in a range of four to 58 articles. Table 7 shows the eight instruments and the criteria and indicators included.

The most frequently used instruments were Flesch-Kincaid readability tests, measuring the readability of web pages. These tests were developed by Kincaid; the core measures include word length and sentence length (Kincaid, Fishburne, Rogers, & Chissom, 1975). The other instrument the SMOG, is also a measure of readability. It estimates the years of education needed to understand a piece of writing (Hedman, 2008).

Instruments used in more than 20 articles included DISCERN, HONcode, and JAMA benchmarks. DISCERN is a brief questionnaire to help health consumers and information providers assess the quality of written information about treatment choices for a health problem (Charnock, 1998). HONcode is a code of ethics that guides site managers in setting up a minimum set of mechanisms to provide quality medical information tailored to users' needs. It consists of eight principles (Health On the Net Foundation, 2013). JAMA benchmarks refer to a set of core standards (authorship, attribution, disclosure, and currency) identified by Silberg et al. (1997) for judging the quality of both print and digital health information. These instruments focus mainly on assessing factors associated with the formality of content, using criteria such as currency and credibility (authoritativeness and trustworthiness). The indicators used in each instrument vary but also overlap. For example, indicators such as the content creation date, the authors' professional credentials, the advertising policy, and sources of content appear in at least two of them. It is worth noting that DISCERN was also used in a number of studies to evaluate the substance of content, specifically the completeness of treatment information. Its indicators include the presence of information concerning multiple treatment choices; how each treatment works; the benefits and risks of each treatment; the effects of the treatment choices on daily activities and relationships with family, friends, and carers; what would happen if no treatment is used; and areas of uncertainty.

One commonality of these instruments was that they paid little attention to design. DISCERN and JAMA benchmarks do not include indicators and criteria explicitly related to site design. HONcode includes a criterion related to accessibility (i.e., the contact of a site must be easy to access from anywhere on the site). The other HONcode criterion related to design was "privacy and data protection," which suggests that sites should have and present a privacy policy regarding how they store users' data (e.g., e-mail, name, or medical data) and how they handle and use the data collected (e.g., for their own statistics or to share with third parties).

The remaining three instruments were LIDA (Minervation validation instrument), Abbott, and Sandvik. LIDA is a validated tool produced by Minervation, a UK-based "spin-out" company from the University of Oxford, which provides a variety of services to help improve the quality of

TABLE 7. Most used pre-existing instruments among the sample articles.

Criteria	Indicator	Flesch-Kincaid readability tests	SMOG	DISCERN	HONCode	JAMA benchmarks	LIDA	Abbott	Sandvik
Completeness	Presence of various aspects concerning treatments			✓					
Currency	Last update time				✓	✓	✓	✓	✓
	Content creation date				✓	✓	✓	✓	✓
Credibility	Author name and professional credentials			✓	✓	✓		✓	✓
	Aim of the site clear				✓	✓	✓	✓	
	Feedback mechanism (contact information)			✓	✓				✓
	Copyright				✓	✓	✓	✓	
	Advertising policy				✓	✓			
	Financial disclosure					✓		✓	
	Interest conflict					✓			
	Owner/sponsor				✓	✓	✓	✓	
	References and source of content			✓		✓	✓	✓	✓
	Additional source of support			✓		✓		✓	✓
	Site type								
	Balanced content			✓				✓	✓
	Target audience			✓			✓		
	Spelling mistakes						✓		
	Editorial process			✓	✓		✓	✓	
Readability		✓	✓				✓	✓	
Aesthetics & Graphic Design	Good layout						✓	✓	
	Relevant images						✓	✓	
	Color scheme						✓	✓	
	Design consistency						✓		
	Headings						✓		
Navigability	Navigation structure & current location						✓	✓	✓
Accessibility	Clear presentation of website contact				✓		✓		✓
	Operational (available, no dead links, and browser-independent)				✓		✓	✓	
	Registration & software requirements						✓		
	Accessible for users with disabilities						✓	✓	
Interactivity	Information exchange (e.g., forums, e-mails)						✓	✓	✓
	Internal search engine						✓		✓
	Multimedia						✓	✓	
	Personalization						✓		
Privacy	Policy on collection and use of personal data				✓				
No. of articles used the tool (%)		58 (35.2)	4 (2.4)	45 (27.3)	29 (17.6)	22 (13.3)	8 (4.8%)	6 (3.6)	4 (2.4)

TABLE 8. Conclusions concerning the overall quality of the sites evaluated.

Overall quality	No. of studies (%)
Negative	91 (55.2%)
Positive	10 (6.1%)
Varied	61 (37.0%)
Neutral	3 (1.8%)
Total	*165*

healthcare information (Tavare, Alsafi, & Hamady, 2012). Unlike the three instruments just mentioned (DISCERN, HONcode, and JAMA benchmarks), LIDA provides a comprehensive set of indicators to measure all but the privacy aspect of the design of a site. Abbott was based on prior literature (Abbott, 2000). In addition to measuring currency, credibility, and readability (using the Flesch Reading Ease Score), it also measures page aesthetics, accessibility, and interactivity. Sandvik was proposed in Sandvik (1999), based partially on criteria suggested by Silberg et al. (1997) and partially on the HONcode principles. As a result, it focuses on the formality of content, but it also evaluates the navigability and interactivity of a health information source.

Evaluation Results

In the selected articles, authors often concluded the evaluation with an overall assessment of information quality. Table 8 shows the distribution of the tones of the conclusions. More than half of the reviewed articles (55.2%) reached an overall negative conclusion concerning information quality. Consistent with the criteria by which quality was evaluated, the reasons cited for the negative conclusions were related to the content of information (e.g., incorrect, incomplete, basic, superficial, and not useful), and to site design (e.g., not appropriate and inaccessible). Ten articles (6.1%) reached an overall positive conclusion. In these cases, the content was characterized as correct or not dangerous, or the design was characterized as attractive, easy to use, and accessible. Sixty-one articles (37.0%) concluded that the quality of health information on the web was highly variable, differing among different sites or site categories (e.g., .gov and .com sites). Only three studies (1.8%) made no explicit conclusive remarks on overall quality, but instead discussed both benefits and risks of the sites.

Discussion

Quality and Evaluation Criteria and Indicators

Consistent with prior reviews (Eysenbach et al., 2002; Kim et al., 1999), this review revealed that quality was defined differently and measured using different combinations of criteria in different studies. Nevertheless, it also makes several contributions to the understanding of quality and its evaluation.

First, a new level of clarity was brought to the concept of quality. Unlike prior studies, we used decision-making theories as a lens to differentiate indicators and criteria. Criteria were defined as high-level abstract rules, and indicators as observable attributes of content or technological platforms. This differentiation enables a two-layered view of quality. The first layer has to do with the content of information on a site and consists of two levels: substance and formality. At the substance level, quality is defined in relation to the accuracy and completeness of the content and is measured against medical textbooks or guidelines. At the formality level, quality is defined in relation to currency, credibility (authoritativeness and trustworthiness), and readability, and is measured by indicators that can be directly observed (e.g., author, copyright, and references). The second layer of quality has to do with the design of a site and consists of two aspects: the technical aspect, including accessibility, aesthetics, navigability, interactivity, and the social aspect, including privacy policy and social and cultural appropriateness.

Second, a comprehensive view of quality indicators was provided. This view suggests potential challenges associated with identifying and locating appropriate indicators in evaluation. An example is that four types of dates were used in the selected articles to evaluate currency: the publication date, the time of the last update, the copyright date, and the date for the next update.

Third, this review revealed a wide use of design-related indicators and criteria in evaluation. Compared to earlier studies, which reported that the design-related criteria mainly used were site layout and navigation (Eysenbach et al., 2002), this review revealed that many studies evaluated design in relation to sites' interactivity (e.g., the enabling of information exchanges among users, social media expansion, the presence of internal search engines, and multimedia) and privacy protection (the presence of a policy on the collection and use of personal data). The attention to interactivity may be attributable to the fast development of eHealth research, an emerging area concerning the design and application of interactive technologies to promote health behavior and management (Hesse & Shneiderman, 2007), and the attention to privacy may be attributable to an increasing concern about data security and patients' privacy (Williams, 2010). Another design-related criterion, social and cultural sensitivity, was also found to be evaluated in three studies (Nasser, Mullan, & Bajorek, 2012; Stinson et al., 2011; Wallace, Turner, Ballard, Keenum, & Weiss, 2005). Although all three studies concluded that sites often failed to cater to users' cultural backgrounds, the use of this criterion, nonetheless, indicates an increased recognition among researchers that health information systems are social-technical systems; high-quality systems should not only be accessible and usable, but also attend to users' cultural differences. This recognition is particularly valuable

as more countries, communities, and organizations around the world are becoming more ethnically diverse (Nasser et al., 2012).

As more attention has been paid to the design of health-related websites, it is worth noting that design may not be a consistent indicator of information quality. For instance, several of the studies reviewed suggested that design factors, specifically aesthetics and page ranks in search engines, were not reliable predictors of a site's overall content quality (Chestnutt, 2002; Kaicker, Wu, & Athreya, 2012; Irwin et al., 2011; Perez-Lopez & Perez Roncero, 2006; Tan, Kostapanagiotou, & Jilaihawi, 2009; Zermatten, Khazaal, Coquard, Chatton, & Bondolfi, 2010). Over the last decade, the design community has devoted considerable effort to making sites appear credible in order to gain users' trust (Fogg, 2003). Design guidelines for creating credible websites have also been developed (Ni Riordain & McCreary, 2009). When the quality of content is not ensured and deliberate efforts go into making sites seem credible, evaluating online health information solely based on design could be problematic.

Quality Evaluation Instruments

Unlike prior reviews, which focused on instruments available on the web, this review focused on those used by researchers to evaluate consumer-oriented health information. Among the most widely used instruments were the Flesch-Kincaid (F-K) readability tests and the SMOG. Although the SMOG was more aligned with the medical field, the F-K tests were adopted more widely. This could be because many authors used MS Word to calculate readability scores, and the program's calculation is based on the F-K tests (MS Office, 2007). The remaining instruments, namely DISCERN, the JAMA benchmarks, the HONcode, LIDA, Abbott, and Sandvik, although differing in their indicators and criteria, tend to focus on measuring factors associated with the formality of content and the design of technological platforms. A close examination reveals that each instrument requires evaluators to examine an average of 13 indicators to reach a judgment. This fact, in keeping with prior studies (Bernstam et al., 2005), suggests that, although these instruments were designed to guide consumers in evaluating health information quality, the sheer amount of effort required to locate and examine this many indicators makes them impractical to use, particularly during information searches when the overriding goal is to find relevant information (Eysenbach & Kohler, 2002; Harland, 2007; Zhang, 2012). The other concern raised involved the validity of some of the popular instruments. A number of the articles in the sample reported conflicting results concerning the ability of DISCERN, HONcode, and JAMA benchmarks to predict health information quality (e.g., Bailey et al., 2012; Barnes, 2009; Hendrick et al., 2012; Tallgren & Backlund, 2009; Tavare et al., 2012; Yeo et al., 2007).

This review also revealed that it was common for researchers to create domain-specific instruments based on medical guidelines, textbooks, literature, and/or evaluators' medical expertise to guide evaluation of the substance, that is, the accuracy and completeness of information content. One third of the reviewed articles constructed such instruments to facilitate the evaluation.

The challenges associated with using the existing instruments, the need to construct specialized instruments for evaluating the substance of content, as well as the aforementioned operational difficulties in identifying appropriate indicators, all could impose great challenges to consumers, suggesting a need for user assistance. Two means can be used to mitigate such challenges. One is user education and training. The skills to critically evaluate online health information can be built into school or library programs to improve consumers' health literacy. Tutorials that teach health information evaluation can be made available on the web to reach a broad audience. The other means is to build tools that can automatically detect quality indicators (Eysenbach & Diepgen, 1998). Such tools may rely on identifying metadata related to the formality of content (e.g., author and publication dates) and to design (e.g., internal search engines, feedback, and user forums) (Wang & Liu, 2007), but at the same time, they should build in domain-specific knowledge to help users evaluate the substance of the content. Some research has taken this direction. For example, Griffiths, Tang, Hawking, and Christensen (2005) developed a tool to automatically rank depression websites according to their evidence-based qualities.

Overall Quality of Health Information Online

Several studies noted an improvement of quality over the years. For example, Zun, Downey, and Brown (2011) found a distinct improvement in the completeness and accuracy of emergency medical information between 2002 and 2008. Nevertheless, like the previous review (Eysenbach et al., 2002), this review revealed a rather gloomy picture, with 55.2% of the articles concluding that the quality of online health information was problematic and 37.0% concluding that quality varied across websites. This finding suggests that consumers should critically evaluate online health information and that efforts should be made to educate consumers and to create effective tools to facilitate quality evaluation.

Directions for Future Studies

Several common methodological shortcomings were acknowledged in many of the selected articles. First, although all the articles were evaluating consumer-oriented health information, the keywords used for searching were mostly not from consumers; rather, they were generated by researchers under the assumption that they were consumers' terms. Second, most studies reviewed a limited number of web pages, mostly top-ranking pages from search engine results lists, under the assumption that they were the most likely to be viewed. Thus, the results represent only a

snapshot of the quality of online health information. Third, in many studies, the review was performed by a single evaluator, which could directly impact the reliability of the results (Craigie, Loader, Burrows, & Muncer, 2002). Future studies on evaluating consumer health information should attempt to mitigate these shortcomings.

Two future research directions are worth pursuing. First, in all the studies included in the review, the evaluators were researchers or medical experts; few involved both domain experts and consumers themselves. Thus, limited knowledge was gained about differences between experts and general consumers in evaluating quality. Future research can fill this gap. Second, several articles in the sample evaluated user-generated content, such as messages in online communities (Hwang et al., 2007), but overall, there was a lack of discussion about the quality of content on social media platforms, such as blogs, Wikipedia, and social networking sites. Most of the articles reviewed did not report whether they had included such content; and for those that did, some included only one or two such sites, while the remainder excluded these types of sites altogether. Thus, this review provided limited knowledge of the quality of health information in emerging new media. As more people began to contribute health information online (Fox, 2011; Hamm et al., 2013), there is an urgent need to understand the quality of user-generated content. Future research should examine questions such as how the evaluation of these new media should differ from that of traditional websites, whether there are relationships between quality and user behavior, and what some potential mechanisms might be for improving the quality of information in these media.

Conclusion

Several findings consistent with prior reviews emerged from this review. First, quality varied across medical domains and across websites, and the overall quality of online health information remains problematic. Second, there was no agreed-upon definition of quality. Different authors operationalized the concept in different ways and evaluated it using a wide variety of indicators and criteria. Third, many instruments were available for measuring quality, but their validity was questionable. The persistence of these problems indicates that fundamental challenges regarding quality evaluation remain and continuing attention needs to be paid to them.

This review also advanced current understanding of quality evaluation. First, by conceptually differentiating quality indicators and criteria, it has made clear that quality has been evaluated in relation to the substance of content, the formality of content, and the design of technological platforms. Second, attention to design elements, particularly interactivity, privacy, and social and cultural appropriateness, is on the rise. This suggests the rapid diffusion of user-centered design principles into the design and evaluation of health information systems, and a growing recognition of the need to study these systems from a social-technical perspective. Third, the existence of a large number of quality indicators and nuanced differences among some of them point to the considerable operational difficulties associated with evaluating health information in the current complex sociotechnical environment. Fourth, there are a number of preexisting instruments, but only a few were used by a handful of the studies, and these instruments tend to focus on evaluating the formality of content. The quality of the substance of content (i.e., accuracy and completeness) was always evaluated with proprietary instruments based on medical guidelines or textbooks. The challenges associated with quality evaluation call for user education, as well as the development of automatic tools. To further advance knowledge of the quality of health information in the fast-changing web environment, future research is needed to study the consumer's perspective, and to explore opportunities offered by emerging new media.

Acknowledgments

References

Abbott, V.P. (2000). Web page quality: can we measure it and what do we find? A report of exploratory findings. Journal of Public Health Medicine, 22(2), 191–197.

Arora, N.K., Hesse, B.W., Rimer, B.K., Viswanath, K., Clayman, M.L., & Croyle, R.T. (2008). Frustrated and confused: The American public rates its cancer-related information-seeking experiences. Journal of General Internal Medicine, 23(3), 223–228.

Bailey, M.A., Coughlin, P.A., Sohrabi, S., Griffin, K.J., Rashid, S.T., Troxler, M.A. . . . (2012). Quality and readability of online patient information for abdominal aortic aneurysms. Journal of Vascular Surgery, 56(1), 21–26.

Barnes, C., Harvey, R., Wilde, A., Hadzi-Pavlovic, D., Wilhelm, K., & Mitchell, P.B. (2009). Review of the quality of information on bipolar disorder on the Internet. Australian & New Zealand Journal of Psychiatry, 43(10), 934–945.

Bernstam, E.V., Sagaram, S., Walji, M., Johnson, C.W., & Meric-Bernstam, F. (2005). Usability of quality measures for online health information: Can commonly used technical quality criteria be reliably assessed? International Journal of Medical Informatics, 74(7–8), 675–683.

Bernstam, E.V., Shelton, D.M., Walji, M., & Meric-Bernstam, F. (2005). Instruments to assess the quality of health information on the World Wide Web: What can our patients actually use? International Journal of Medical Informatics, 74(1), 13–19.

Brunswik, E. (1952). The conceptual framework of psychology. Chicago: University of Chicago Press.

Car, J., Lang, B., Colledge, A., Ung, C., & Majeed, A. (2011). Interventions for enhancing consumers' online health literacy. The Cochrane Database of Systematic Reviews 2011, (6). Article No. CD007092.

Charnock, D. (1998). THE DISCERN handbook: Quality criteria for consumer health information on treatment choices. Retrieved from: http://www.discern.org.uk/discern.pdf

Chestnutt, I.G. (2002). The nature and quality of periodontal related patient information on the world-wide web. British Dental Journal, 193(11), 657–659; discussion 643.

Chou, W., Prestin, A., Lyons, C., & Wen, K. (2013). Web 2.0 for health promotion: Reviewing the current evidence. American Journal of Public Health, 103(1), e9–18.

Cline, R.J.W., & Haynes, K.M. (2001). Consumer health information seeking on the Internet: the state of the art. Health Education Research, 16(6), 671–692.

Craigie, M., Loader, B., Burrows, R., & Muncer, S. (2002). Reliability of health information on the Internet: an examination of experts' ratings. Journal of Medical Internet Research, 4(1), e2. Available at: http://www.jmir.org/2002/1/e2/

Crocco, A.G., Villasis-Keever, M., & Jadad, A.R. (2002). Analysis of cases of harm associated with use of health information on the internet. JAMA, 287(21), 2869–2871.

Eysenbach, G. (2008). Medicine 2.0: Social networking, collaboration, participation, apomediation, and openness. Journal of Medical Internet Research, 10(3). Available at: http://www.jmir.org/2008/3/e22/

Eysenbach, G., & Diepgen, T.L. (1998). Towards quality management of medical information on the internet: evaluation, labelling, and filtering of information. BMJ, 317(7171), 1496–1502.

Eysenbach, G., & Kohler, C. (2002). 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, 324(7337), 573–577.

Eysenbach, G., Powell, J., Kuss, O., & Sa, E.-R. (2002). Empirical studies assessing the quality of health information for consumers on the World Wide Web. JAMA, 287(20), 2691–2700.

Feufel, M.A., & Stahl, S.F. (2012). What do Web-use skill differences imply for online health information searches? Journal of Medical Internet Research, 14(3), e87. Available at: http://www.jmir.org/2012/3/e87/

Fogg, B.J. (2003). Persuasive technology: Using computers to change what we think and do. San Franciso: Morgan Kaufmann.

Fox, S. (2011). The social life of health information. Retrieved from: http://www.pewinternet.org/files/old-media/Files/Reports/2011/PIP_Social_Life_of_Health_Info.pdf

Gagliardi, A., & Jadad, A.R. (2002). Examination of instruments used to rate quality of health information on the internet: Chronicle of a voyage with an unclear destination. BMJ, 324(7337), 569–573.

Griffiths, K.M., Tang, T.T., Hawking, D., & Christensen, H. (2005). Automated assessment of the quality of depression Websites. Journal of Medical Internet Research, 7(5), e59. Available at: http://www.jmir.org/2005/5/e59/

Hamm, M.P., Chisholm, A., Shulhan, J., Milne, A., Scott, S.D., Given, L.M. . . . (2013). Social media use among patients and caregivers: A scoping review. BMJ open, 3(5), e002819.

Harland, J.P. (2007). Assessing the quality of websites providing information on multiple sclerosis: Evaluating tools and comparing sites. Health Informatics Journal, 13(3), 207–221.

Hesse, B.W., & Shneiderman, B. (2007). eHealth research from the user's perspective. American Journal of Preventive Medicine, 32(5 Suppl), S97–103.

Health On the Net Foundation. (2013). The HON Code of Conduct for medical and health web sites (HONcode). Accessed from: http://www.hon.ch/HONcode/Conduct.html

Hedman, A.S. (2008). Using the SMOG formula to revise a health-related document. American Journal of Health Education, 39(1), 61–64.

Hendrick, P.A., Ahmed, O.H., Bankier, S.S., Chan, T.J., Crawford, S.A., Ryder, C. R. . . . (2012). Acute low back pain information online: An evaluation of quality, content accuracy and readability of related websites. Manual Therapy, 17(4), 318–324.

Hogarth, R.M. (1987). Judgement and choice (2 ed.). New York: John Wiley & Sons.

Hwang, K.O., Farheen, K., Johnson, C.W., Thomas, E.J., Barnes, A. S., & Bernstam, E.V. (2007). Quality of weight loss advice on Internet forums. The American Journal of Medicine, 120(7), 604–609.

Irwin, J.Y., Thyvalikakath, T., Spallek, H., Wali, T., Kerr, A. R., & Schleyer, T. (2011). English and Spanish oral cancer information on the Internet: A pilot surface quality and content evaluation of oral cancer Web sites. Journal of Public Health Dentistry, 71(2), 106–116.

Jadad, A.R., & Gagliardi, A. (1998). Rating health information on the Internet: Navigating to knowledge or to Babel? JAMA, 279(8), 611–614.

Kaicker, J., Wu, K., & Athreya, S. (2012). Assessing the reliability and quality of online uterine fibroid embolization resources. CardioVascular and Interventional Radiology, 36(2), 385–394.

Khazaal, Y., Chatton, A., Cochand, S., Coquard, O., Fernandez, S., Khan, R. . . . (2010). Quality of web-based information on alcohol dependence. Drugs: Education, Prevention, and Policy, 17(3), 248–260.

Kim, P., Eng, T.R., Deering, M.J., & Maxfield, A. (1999). Published criteria for evaluating health related web sites: Review. BMJ, 318(7184), 647–649.

Kincaid, J., Fishburne, R.P., Rogers, R.L., & Chissom, B.S. (1975). Derivation of new readability formulas (Automated Readability Index, Fog Count, and Flesch Reading Ease formula) for Navy enlisted personnel. Research Branch Report 8-75. Chief of Naval Technical Training: Naval Air Station, Memphis, TN.

Lam, C.G., Roter, D.L., & Cohen, K.J. (2013). Survey of quality, readability, and social reach of websites on osteosarcoma in adolescents. Patient Education and Counseling, 90(1), 82–87.

Meric, F., Bernstam, E.V., Mirza, N.Q., Hunt, K.K., Ames, F.C., Ross, M.I. . . . (2002). Breast cancer on the world wide web: cross sectional survey of quality of information and popularity of websites. BMJ, 324(7337), 577–581.

Mousiolis, A., Michala, L., & Antsaklis, A. (2012). Polycystic ovary syndrome: double click and right check. What do patients learn from the Internet about PCOS? European Journal of Obstetrics & Gynecology and Reproductive Biology, 163(1), 43–46.

Nasser, S., Mullan, J., & Bajorek, B. (2012). Assessing the quality, suitability and readability of internet-based health information about warfarin for patients. Australasian Medical Journal, 5(3), 194–203.

Ni Riordain, R., & McCreary, C. (2009). Head and neck cancer information on the internet: Type, accuracy and content. Oral Oncology, 45(8), 675–677.

O'Grady, L. (2006). Future directions for depicting credibility in health care web sites. International Journal of Medical Informatics, 75(1), 58–65.

O'Keefe, D.J. (2002). Persuasion: Theory and research (2nd ed.). Thousand Oaks, CA: Sage Publications.

Perez-Lopez, F.R., & Roncero, G.R.P. (2006). Assessing the content and quality of information on the treatment of postmenopausal osteoporosis on the World Wide Web. Gynecological Endocrinology, 22(12), 669–675.

Pew Research Center. (December 16, 2013). Health fact sheet. Accessed from: http://www.pewinternet.org/fact-sheets/health-fact-sheet/

Post, R.E., & Mainous, A.G., 3rd. (2010). The accuracy of nutrition information on the Internet for type 2 diabetes. Archives of Internal Medicine, 170(16), 1504–1506.

Risk, A., & Dzenowagis, J. (2001). Review of Internet health information quality initiatives. Journal of Medical Internet Research, 3(4), e28. Available at: http://www.jmir.org/2001/4/e28/

Sandvik, H. (1999). Health information and interaction on the internet: a survey of female urinary incontinence. BMJ (Clinical Research ed.), 319(7201), 29–32.

Silberg, W.M., Lundberg, G.D., & Musacchio, R.A. (1997). Assessing, controlling, and assuring the quality of medical information on the Internet: Caveant lector et viewor—Let the reader and viewer beware. JAMA, 277(15), 1244–1245.

Stanovich, K.E. (1999). Who is rational?: Studies of individual differences in reasoning. New York: Psychology Press.

Stinson, J.N., White, M., Breakey, V., Chong, A.L., Mak, I, Low, K.K. . . . (2011). Perspectives on quality and content of information on the internet for adolescents with cancer. Pediatric Blood & Cancer, 57(1), 97–104.

Tallgren, M., & Backlund, M. (2009). Patient information about general anaesthesia on the internet. Anaesthesia, 64(4), 408–415.

Tan, B.H.L., Kostapanagiotou, K., & Jilaihawi, A.N.A. (2009). A review of mesothelioma information on the World Wide Web. Journal of Thoracic Oncology, 4(1), 102–104.

Tavare, A.N., Alsafi, A., & Hamady, M.S. (2012). Analysis of the quality of information obtained about uterine artery embolization from the Internet. Cardiovascular and Interventional Radiology, 35(6), 1355–1362.

Tseng, S., & Fogg, B.J. (1999). Credibility and computing technology. Communications of the ACM, 42(5), 39–44.

Tsunehara, S.T., & Gaster, B. (2002). A critical assessment of diabetes information on the Internet. Diabetes Educator, 28(2), 186.

Wallace, L.S., Turner, L.W., Ballard, J.E., Keenum, A.J., & Weiss, B.D. (2005). Evaluation of web-based osteoporosis educational materials. Journal of Women's Health, 14(10), 936–945.

Wang, Y.L., & Liu, Z.K. (2007). Automatic detecting indicators for quality of health information on the Web. International Journal of Medical Informatics, 76(8), 575–582.

Williams, J. (2010). Social networking applications in health care: Threats to the privacy and security of health information. In Proceedings of the 2010 ICSE Workshop on Software Engineering in Health Care (SEHC), pp. 39–49.

Wilson, P. (2002). How to find the good and avoid the bad or ugly: A short guide to tools for rating quality of health information on the internet. BMJ, 324(7337), 598–602.

Wyatt, J.C. (1997). Commentary: Measuring quality and impact of the world wide web. BMJ, 314(7098), 1879–1879.

Yeo, H., Roman, S., Air, M., Maser, C., Trapasso, T., Kinder, B. . . . (2007). Filling a void: Thyroid cancer surgery information on the internet. World Journal of Surgery, 31(6), 1185–1191.

Zermatten, A., Khazaal, Y., Coquard, O., Chatton, A., & Bondolfi, G. (2010). Quality of web-based information on depression. Depression and Anxiety, 27(9), 852–858.

Zhang, Y. (2012). Consumer health information searching process in real life settings. In Proceedings of the 75th Annual Conference of the American Society for Information Science & Technology (ASIST'12), pp. 1–10.

Zhang, Y. (2014). Beyond quality and accessibility: Source selection in consumer health information searching. Journal of the Association for Information Science and Technology, 65(5), 911–927.

Zun, L.S., Downey, L., & Brown, S. (2011). Completeness and accuracy of emergency medical information on the Web: Update 2008. Western Journal of Emergency Medicine, 12(4), 448–454.