The Effect of Search Term on the Quality and Accuracy of Online Information Regarding Distal Radius Fractures

Christopher J. Dy, MD, MSPH, Samuel A. Taylor, MD, Ronak M. Patel, BS, Alison Kitay, MD, Timothy R. Roberts, MLS, Aaron Daluiski, MD

Purpose  Recent emphasis on shared decision making and patient-centered research has increased the importance of patient education and health literacy. The internet is rapidly growing as a source of self-education for patients. However, concern exists over the quality, accuracy, and readability of the information. Our objective was to determine whether the quality, accuracy, and readability of information online about distal radius fractures vary with the search term.

Methods  This was a prospective evaluation of 3 search engines using 3 different search terms of varying sophistication (“distal radius fracture,” “wrist fracture,” and “broken wrist”). We evaluated 70 unique Web sites for quality, accuracy, and readability. We used comparative statistics to determine whether the search term affected the quality, accuracy, and readability of the Web sites found. Three orthopedic surgeons independently gauged quality and accuracy of information using a set of predetermined scoring criteria. We evaluated the readability of the Web site using the Fleisch-Kincaid score for reading grade level.

Results  There were significant differences in the quality, accuracy, and readability of information found, depending on the search term. We found higher quality and accuracy resulted from the search term “distal radius fracture,” particularly compared with Web sites resulting from the term “broken wrist.” The reading level was higher than recommended in 65 of the 70 Web sites and was significantly higher when searching with “distal radius fracture” than “wrist fracture” or “broken wrist.” There was no correlation between Web site reading level and quality or accuracy.

Conclusions  The readability of information about distal radius fractures in most Web sites was higher than the recommended reading level for the general public. The quality and accuracy of the information found significantly varied with the sophistication of the search term used.

Clinical relevance  Physicians, professional societies, and search engines should consider efforts to improve internet access to high-quality information at an understandable level. (J Hand Surg 2012;xx:. Copyright © 2012 by the American Society for Surgery of the Hand. All rights reserved.)

Key words  Distal radius, education, health literacy, internet, readability.
Patients are increasingly using the internet to access health care information. The internet’s usability and versatility allow patients to inform themselves about medical diagnoses and treatments in an unprecedented manner. Many health care providers are modifying their practices to incorporate online information in their patient interactions. However, concern exists regarding the quality, accuracy, and readability of the information available. As shared decision making and patient-centered outcomes research occupy a larger role in the public health agenda, ensuring access to accurate and understandable health information online will become increasingly important.

Prior studies across multiple medical disciplines, including orthopedic surgery, have described poor-quality information that is published on the internet, sometimes with secondary profit motives. Furthermore, patient materials produced by physician specialty organizations, including surgical subspecialties and primary care disciplines, are often written at a level that may not be readable and understandable by the general public. Because misinformation could potentially influence a patient’s medical decision making, physicians and the public must be aware of the quality, accuracy, and readability of online medical information.

Disparate access to medical information comes in many forms. There are well-described differences in access to computers, internet information, and reading level–appropriate health care literature. Because access to online health care information is largely filtered by the major search engines, we asked whether the health care information retrieved was affected by the search term used. In the current study, we investigated the quality, accuracy, and readability of information available on the internet regarding distal radius fractures (DRFs). We chose this subject for investigation because DRFs are common acute musculoskeletal injuries that are often associated with or predisposed by osteoporosis, which is responsible for approximately 2 million fractures annually, and which is projected to have an annual direct cost of $25.3 billion by 2025. We used 3 different search terms in 3 different search engines to capture the variability in search strategies, hypothesizing that the readability of information, but not the quality or accuracy, available on the internet about DRFs would vary with the search term used.

MATERIALS AND METHODS

We selected the search terms “distal radius fracture,” “wrist fracture,” and “broken wrist” to simulate the variability of search terms used when seeking information about DRFs. Because it has been shown that patients are generally unsure about musculoskeletal anatomy and orthopedic terminology (such as “fracture” and “reduction”), we included the words “wrist” and “broken” in our search terms.

We entered each of the 3 search terms into Google, Yahoo, and Bing on July 24, 2011, within a single session for a total of 9 separate searches. We selected these search engines because they represent approximately 93% of internet searches performed. We compiled the first 25 results from each search and eliminated duplicate results and nonfunctional Web sites, which left a list of 107 unique Web sites (Fig. 1). We accessed all of the Web sites during a 2-hour period and created an electronic capture of each Web site after excluding 23 Web sites with only news items (often related to injuries to professional athletes). We also excluded sites from further review if they specifically addressed non-DRF wrist injuries (6 sites) such as carpal fractures or contained materials explicitly intended for peer review (8 sites). Sponsored results were not prioritized on the search result list. A total of 70 unique Web sites remained for final review (Fig. 1).

We assessed the quality and accuracy of the information on the Web sites in a manner similar to prior investigations of information about scoliosis and disc herniation. We generated a content quality score that included 30 items related to the pathophysiology, evaluation, and treatment of distal radius fracture (Table 1). The 30 items in the content quality score represent what should be presented to patients if they are seeking information about DRF on the internet. We included the 3 guidelines published by the American Academy of Orthopedic Surgeons that met the working group’s definition for a strong evidence-based recommendation. We derived the remaining 27 items from the American Society for Surgery of the Hand’s patient information Web site about DRF. Similar to a prior investigation, we reviewed the Web site quality and awarded 1 point when a Web site contained correct information for each item, with a maximum score of 30. Three independent reviewers (all PGY4 orthopedic surgery residents, each trained by the senior author in the diagnosis, evaluation, and treatment of DRF) evaluated the quality of each Web site using an identical electronic capture of the Web site. We averaged the scores of the 3 reviewers to provide a mean score for each Web site.

To assess Web site accuracy, the 3 reviewers independently rated the accuracy of information on the Web site on a scale of 1 to 4. An accuracy score of 1 represented agreement with less than 25% of the information on the Web site; 2 represented agreement with
We evaluated the readability of each Web site using the Fleisch-Kincaid (FK) method of analysis, which has previously been used when evaluating information about orthopedic and upper extremity conditions. After preparing the text identically to that in Wang and colleagues, we used Microsoft Word (Redmond, WA) to determine the FK readability grade level of each Web site. The FK grade level indicates that a person who has completed that academic grade level will be able to read and comprehend the material. A higher FK grade level is assigned to material that is more difficult to read and understand.

We grouped the Web sites by the search term used to find them: “distal radius fracture,” “wrist fracture,” or “broken wrist.” When a Web site was retrieved using more than 1 search term, we categorized the Web site under the search term that yielded the earliest result. For example, when a Web site was the third search result under “distal radius fracture” but was the 10th result under “wrist fracture,” the Web site was categorized with the “distal radius fracture” group. We also grouped the Web sites by the highest-priority result (“hits”): 1–5, 6–10, 11–15, 16–20, and 21–25. We grouped Web sites into those with an FK score above the sixth-grade level and those at or below the sixth-grade level, in accordance with prior recommendations for patient education materials. In addition, we noted whether the Web sites had a potential commercial bias (contained advertisements for hand, wrist, or trauma-related products or services). Finally, we categorized the Web sites by authorship: health care provider (physician, nurse, or physical therapist with explicitly stated credentials), commercial site authored by non–health care provider; unspecified non–health care provider with noncommercial Web sites (eg, message boards, blogs, and personal Web sites), or physician specialty society.

We calculated descriptive statistics for quality score, summed accuracy assessment, and FK level. We evaluated normality of the data using skewness and kurtosis. We analyzed non-normally distributed data using nonparametric tests and used analysis of variance (for normally distributed data) and Kruskal-Wallis test (for non-normally distributed data) with post hoc pairwise comparisons to determine differences in quality, accuracy, and readability based on the search term used, the order of search results, and the Web site author. We used parametric statistical tests to compare quality and FK scores (both normally distributed), whereas we used non-parametric tests to compare accuracy (non-normally distributed). We employed independent sample t-tests (for normally distributed data) or Mann-Whitney U tests (for non-normally distributed data) to determine differences in quality, accuracy, or readability, based on whether a Web site had a potential com-
multivariate regression models to determine whether Web site quality or accuracy was significantly influenced by search term used, while controlling for the Web site’s FK score. The threshold for statistical significance was $P < .050$ in all statistical tests.

**RESULTS**

Of the 70 Web sites, quality (12.0 ± 7.2 [mean ± SD] of a maximum score of 30; range, 0–26) and accuracy (11.1 ± 2.2 of a maximum score of 12; range, 3–12) varied greatly. The average FK level was 10.0 ± 2.3 (range, 4.4–15.0); only 5 (8%) of the 70 Web sites had the recommended FK score of sixth grade or lower. Data for quality and FK level were normally distributed, whereas data for accuracy were not.

Of the 70 total Web sites, 28 were identified using the search term “distal radius fracture,” 21 with “wrist fracture,” and 21 with “broken wrist.” We categorized 19 (27%) as seeking commercial gain. A total of 20 Web sites were authored by health care providers, 40 were written by non–health care providers for commercial Web sites, 5 were written by non–physician health care providers for noncommercial Web sites, and 5 were written by physician specialty societies.

There was no difference in quality or accuracy based on a readability threshold of sixth grade (higher than sixth grade vs lower than sixth grade). However, when using a threshold of eighth-grade reading level, there was a difference in quality ($P = .034$) but not accuracy. There was no correlation between FK grade level and quality or between FK grade level and accuracy.

The FK readability scores varied significantly between search term groups ($P < .001$, analysis of variance). Post hoc pairwise comparisons demonstrated a higher FK score in Web sites found with the term “distal radius fracture” than in Web sites found with the terms “wrist fracture” ($P = .013$) and “broken wrist” ($P < .001$) (Table 2).

Accuracy of the content was different between the search term groups ($P = .008$, Kruskal-Wallis test). Pairwise comparisons showed higher accuracy in Web sites found with the search term “distal radius fracture” than with the term “broken wrist” ($P = .006$) (Table 2).

There was a difference in quality when we compared the “distal radius,” “wrist fracture,” and “broken wrist” groups ($P = .048$, analysis of variance). Pairwise comparisons demonstrated a difference between the “distal radius fracture” and “broken wrist” groups ($P = .019$) (Table 2). We constructed 2 separate multivariate linear regression models to evaluate the influence of search term on quality and accuracy while controlling for the FK score.
TABLE 2. Quality, Accuracy, and FK Grade Level Results for Web Sites, Categorized by Search Term Used

<table>
<thead>
<tr>
<th>Search Term</th>
<th>Quality (Maximum 30)</th>
<th>Overall P Value</th>
<th>Accuracy (Maximum 12)</th>
<th>Overall P Value</th>
<th>FK Grade</th>
<th>Overall P Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>“Distal radius fracture” (n = 70)</td>
<td>14.0 ± 8.2b</td>
<td>.048</td>
<td>11.9 ± 0.4c</td>
<td>.008</td>
<td>11.2 ± 2.1de</td>
<td>&lt; .001</td>
</tr>
<tr>
<td>“Wrist fracture” (n = 28)</td>
<td>13.7 ± 8.3</td>
<td>ANOVA</td>
<td>10.1 ± 3.1</td>
<td>Kruskal-Wallisa</td>
<td>9.7 ± 2.3d</td>
<td>ANOVA</td>
</tr>
<tr>
<td>“Broken wrist” (n = 21)</td>
<td>8.8 ± 5.9b</td>
<td></td>
<td>11.2 ± 2.1c</td>
<td>.075</td>
<td>8.7 ± 1.9e</td>
<td></td>
</tr>
</tbody>
</table>

*Used because of nonparametric distribution of accuracy data.

bStatistically significant pairwise comparison (P = .019).

cStatistically significant pairwise comparison (P = .006).

dStatistically significant pairwise comparison (P = .013).

eStatistically significant pairwise comparison (P < .001).

of each Web site. When controlling for FK level, search term did not affect Web site quality (P = .056; β = −0.256). However, search term affected Web site accuracy (P = .028; β = −0.285), even when controlling for FK level.

There was a difference in quality when we compared Web sites by authorship (P = .002, analysis of variance); pairwise comparisons showed lower quality on personal Web sites (4.2 ± 3.6) compared with Web sites authored by health care providers (13.0 ± 7.2; P = .017) and commercial Web sites (11.8 ± 7.6; P = .30). We did not conduct comparisons of quality with specialty society Web sites because the quality content metric was designed using the American Society for Surgery of the Hand’s patient information Web site (which was 1 of the Web sites evaluated).

The accuracy was different between the authorship groups (P < .001, Kruskal-Wallis test); the post hoc pairwise comparisons showed a significantly lower accuracy on personal Web sites (5.2 ± 2.7) compared with Web sites authored by health care providers (12.0 ± 0.2; P < .001), commercial Web sites (11.4 ± 1.6; P < .001), or specialty societies (12.0 ± 0.0; P < .001). There was no difference in accuracy among the 3 other categories of authors. There was no difference in FK score regardless of who authored the site.

Web sites with a potential commercial bias had lower quality (8.3 ± 5.4, n = 19 vs 13.8 ± 8.2, n = 51; P = .005) and lower accuracy (10.5 ± 2.3 vs 11.4 ± 2.1; P = .002) compared with those without potential commercial bias. There was no difference in FK score based on the potential for commercial gain.

There was no difference in quality, accuracy, and FK score when we compared the Web sites by the order of search results (hits: 1–5, 6–10, 11–15, 16–20, and 21–25). There was no difference in quality, accuracy, or FK score when we compared the Web sites in the first 5 hits with all other Web sites.

**DISCUSSION**

Although patients are increasingly turning to the internet for health care information, we demonstrated a wide level of variation in quality, accuracy, and readability across a spectrum of acute and chronic diseases.15–23 In the current study, we showed that these variations in quality, accuracy, and readability are driven by a number of factors, including the search term used, the Web site’s author, and the potential for commercial bias.

We found lower-quality and less accurate information on the internet using the less sophisticated search term “broken wrist” versus the medical term “distal radius fracture.” The lack of accurate information associated with less sophisticated search terms is problematic, because many patients are generally unsure of proper medical vocabulary. This is particularly problematic for orthopedic terminology.26 For example, even highly educated patients are often unaware that a fracture is the same as a break. Patients across all socioeconomic groups will be less likely to use a search term that will lead to a source of accurate and high-quality information.

Information retrieved with the less sophisticated search term “broken wrist” remained above the recommended reading grade level for health care information (sixth grade) and the average reading level for the United States (eighth grade). Patients who are ultimately able to access Web sites of higher quality (such as those found using the search term “distal radius fracture”) ultimately encounter content that is substantially more difficult to interpret, with a mean FK grade level of 11.2.

The readability level of a Web site affects the reader’s ability to understand the information that is pre-
sented. Only 5 of the 70 Web sites that we evaluated met the reading level of sixth grade recommended by the National Institutes of Health and American Medical Association. Although Web sites written at a higher reading level may be useful for non-orthopedic physicians to help in educating patients, there is a clear need for more information about DRF on the internet that can be read and understood easily by the general population. Physicians organizations or specialty societies may be the most appropriate resources for sound medical information that is also appropriately understandable for all Web site visitors, because secondary profit motive is less likely than commercial Web sites, medical industry, medical centers, or individual physicians. For example, the American Society for Surgery of the Hand is making a concerted effort to ensure that its educational materials are accurate and accessible (Lee SK, personal communication). In addition, both the American Academy of Orthopedic Surgeons and the American Academy of Pediatrics have created online educational portals with information specifically designed for patients and approved by their members. Furthermore, search engines should consider efforts to improve their search algorithms to ensure that high-quality information is made available via nearly synonymous terms such as “broken wrist” and “distal radius fracture.”

Our findings are limited by the lack of a previously established standardized method to review the quality and accuracy of information about DRF. Because of this limitation, we assessed the quality of each Web site relative to the information presented by the American Society for Surgery of the Hand patient information webpage and the American Academy of Orthopedic Surgeons clinical guidelines. We believe that these are the best publicly available sources of information. In addition, we chose our search terms arbitrarily, and it is possible that other search terms may have yielded other results. However, we feel that the search terms we have chosen reflect the vocabulary used by patients to describe the same injury.

Decreased health literacy (lower ability to obtain, process, and understand health care information) is associated with poorer health outcomes and poorer use of health care services. Physicians should continue to educate their patients during face-to-face office visits, but they should also consider integration of preselected internet resources as an opportunity to extend the educational experience beyond the office visit at a pace controlled by the patient. However, we have shown that the quality and accuracy of the content about DRF is driven by something as innocuous as the search term used. Patients who are unfamiliar with more advanced medical terminology are at risk for finding information that is of lower quality and accuracy and is also difficult to read or interpret. When compounded by the lower likelihood of internet access in patients with lower education levels, the inequality of information found with less sophisticated search terms creates a scenario in which reliable health care information is least accessible to those who may need it the most.

REFERENCES


