Assessing the Quality, Reliability, and Readability of Online Information on Dry Eye Disease

Marko Oydanich, MS,* Eric Kuklinski, BA,* and Penny A. Asbell, MD, FACS, MBA†

**Purpose:** The purpose of this study was to assess the quality, reliability, readability, and technical quality of web sites relating to dry eye disease.

**Methods:** A cross-sectional study was conducted that evaluated the first 75 web sites on a Google Search by using the keyword “dry eyes.” Each web site was evaluated by 2 independent reviewers using the DISCERN, HONcode, and JAMA criteria to assess quality and reliability. Interrater reliability was also analyzed. Readability was measured using the Flesch–Kincaid readability tests and the Gunning fog, Simple Measure of Gobbledygook, Coleman–Liau, and automated readability indices. Technical quality was determined by the presence of 10 specific features. Web sites were further categorized into institutional (academic centers, medical associations, and government institutions) and private (private practices) categories.

**Results:** There was no significant difference in scoring observed between the 2 reviewers. The overall mean DISCERN score ± standard error (SE) was 3.2 ± 0.1, the mean HONcode score (±SE) was 9.3 ± 0.3, and the mean JAMA score (±SE) was 1.9 ± 0.1. Institutional web sites had a higher DISCERN score (3.4 ± 0.1 vs. 3.1 ± 0.1; \( P < 0.05 \)) and HONcode score (10.3 ± 0.5 vs. 8.8 ± 0.4; \( P < 0.05 \)) than private sites. Technical quality was higher in institutional web sites compared with private web sites (\( P < 0.05 \)). Readability was poor among all web sites, with most web sites not achieving below a ninth grade reading level.

**Conclusions:** Quality, reliability, and readability scores were low for most web sites. Although institutional web sites achieved higher scores than private web sites, revision is warranted to improve their overall quality of information and readability profile.

---

**Key Words:** dry eye disease, web sites, quality, reliability, readability

(Cornea 2022;41:1023–1028)

The advent of the internet created a means of easy access to information by the public, including health care information. Only approximately 7% of adults do not use the internet,¹ and most users of the internet will use it as a means of learning more about their health conditions and possible treatments.² Many use this information to help make health decisions³ and will consult the internet before consulting their own physician.⁴ However, most information on the internet is not regulated, and most search requests list results according to popularity rather than quality or trustworthiness.⁵ With the high use of the internet for health information and many people trusting the information they find on the internet,⁶ it is important to determine whether the information presented is appropriate and reliable. Several criteria have been developed to analyze the quality and reliability of information such as the DISCERN,⁷ HONcode,⁸ and JAMA benchmark⁹ criteria questionnaires. However, although some web sites may have already been evaluated and certified using these criteria, most commonly the HONcode,⁵ many web sites have not, and therefore, patients may fall victim to consuming unaccredited or even false health information.

Furthermore, the information presented on these web sites for patients may be too complicated for some to understand. The average reading level of an American adult is between the seventh and eighth grade.¹⁰,¹¹ The American Medical Association recommends that the reading difficulty of informational material for patients should be no higher than a sixth-grade level,¹¹ but yet the presentation of health information for patients on many informational web sites is well above that level.¹² This leads to a considerable lack of understanding of certain diseases among patients, which can lead to serious consequences, including a greater risk of hospitalization and higher health care costs, making readability a crucial metric of patient information.¹³ There have been several studies that have assessed the quality, reliability, and readability of online information in many fields including plastic surgery, neurosurgery, and ophthalmology.¹²,¹⁴–¹⁷ However, there has yet to be an investigation of the online information relating to dry eye disease (DED). DED is a complex, multifactorial disease with many possible etiologies which affects approximately 14% of adults in the United States and is one of the most common...
reasons that patients seek eye care. In addition, DED has a substantial impact on the economy, costing approximately $55 billion to the US economy annually and significantly affecting a person’s work productivity. Patients attempting to diagnose themselves or learn more about DED through internet resources may have trouble connecting their signs, symptoms, and treatment options with the information provided on the internet. This is especially apparent in DED, a disease whose treatments are constantly evolving, as we better understand the multifactorial mechanisms leading to its manifestation. Therefore, this study aims to assess the quality, reliability, readability, and technical qualities of web sites that provide information aimed at patients suffering from DED.

**MATERIALS AND METHODS**

A retrospective study on web sites pertaining to DED was conducted. This study did not require approval from the institutional review board because it was observational in nature and did not involve human subjects.

**Search Methodology**

During the month of June 2021, an analysis of 75 web sites relating to DED was conducted using Google.com as the primary search engine. The keyword “dry eye” was used. Web sites were analyzed in the order in which they appeared after the search. The only criterion for inclusion was web sites that were dedicated to informing patients about the definition, causes, symptoms, treatment, and prevention of DED. Web sites that appeared as targeted ads when searched, web sites that were not in English, and web sites not focused on humans were excluded from this study. Web sites were then categorized into either institutional or private sources. Institutional sources were defined as those originating from academic centers, medical associations, or government institutions, while private sources were defined as those originating from private practices.

**Web site Assessment and Analysis**

Each web site underwent an evaluation of its quality, reliability, technical quality, and readability. Two reviewers assessed each web site independently and were evaluated at similar times to avoid any web site updates or changes. Interrater reliability was assessed by comparing each reviewer’s score for every web site analyzed. If scoring of a web site deviated by more than a 10% margin, a third reviewer assessed the web site. The mean of the reviewers’ scores was determined and used as the final overall score for that web site. To assess quality and reliability, 3 separate criteria were used: DISCERN, Health on the Net Code (HONcode), and JAMA criteria. Each criterion has a different set of questions which is meant to evaluate each web site for the quality of information it produces and its comprehensiveness on the topic of interest. The DISCERN criteria are a set of 16 questions, each of which is scored from 1 to 5. The final score is calculated by averaging all the scores, with the best score being 5 and the worst score being 1. The HONcode criteria, developed in 1998 by the Health on the Net Foundation, are a set of 8 questions, each of which is scored from 0 to 2. Each question asks whether a certain principle is fulfilled by the web site. The principles are as follows: authoritativeness, complementarity, privacy, attribution, justifiability, transparency, financial disclosure, and sponsorship. If a score of 2 is given, it means the principle is met completely; if a score of 1 is given, then the principle is met somewhat; and if a score of 0 is given, then the principle is not met. The final score is calculated using the sum of each question’s score resulting in the best score being 16 and the worst score being 0. Finally, the JAMA criteria were also used to assess quality and reliability. It is composed of a set of 4 qualities, that is, authorships, attribution, ownership, and currency, all of which are scored as 0 (does not meet the criteria) or 1 (meets the criteria). The scores are added to receive a final score where 4 is the best score and 0 is the worst score.

Technical quality was assessed based on the following 10 criteria: the absence of ad pop-ups, the absence of ads in the text, does the web page load in 3 seconds or less, is a mobile site available, can the web page be viewed in multiple browsers, the presence of links to social media (eg, Facebook and Twitter), was the web page up to date (ie, updated within 1 year), link to a blog page provided, the page is confidential and does not provide any patient information, and the presence of contact information. If the answer was yes, the page received a score of 1 for that respective criterion; if no, it received a 0. An overall score for technical quality was then determined using the average of all the scores from 0 to 1, 1 being the highest score.

Readability was evaluated using 6 different readability measures, including the Flesch–Kincaid reading ease score, the Flesch–Kincaid grading level, the Gunning fog index, the Simple Measure of Gobbledygook (SMOG) index, the Coleman–Liau index, and the automated readability index. Each measure was determined using the freely accessible online readability tool (webfx.com/tools/readable/). The results from each measure were compared with the average reading level of an American, that is, between the seventh and eighth grade reading level.

**Statistical Analysis**

A Student t test was used to compare overall scores for DISCERN, HONcode, JAMA, and technical quality among both reviewers. Statistical comparison of the DISCERN, HONcode, JAMA, technical quality, and readability scores between institutional and private web sites was also compared using a Student t test. Statistical tests were performed using GraphPad Prism (version 5.1). All data are presented as mean ± standard error (SE).

**RESULTS**

All 75 web sites are evaluated using the DISCERN, HONcode, and JAMA criteria. The overall mean DISCERN score (±SE) was 3.2 ± 0.1 (range: 1–5), the mean HONcode
score (±SE) was 9.3 ± 0.3 (range: 0–16), and the mean JAMA score (±SE) was 1.9 ± 0.1 (range: 0–4). Each reviewer also evaluated each web site’s technical quality independently. The overall mean technical quality score (±SE) was 0.8 ± 0.1 (range: 0–1). All 3 quality/reliability criteria scores and the overall technical quality score did not show any significant difference (P > 0.05) between the 2 reviewers (Table 1).

Readability of all 75 web sites was assessed, and the data were plotted to observe the overall distribution of scores for each measure (Fig. 1). Most of the web sites, 35 (47%), had a Flesch–Kincaid reading ease score between 51 and 60 (Fig. 1A). However, only a total of 18 web sites (24%) had a Flesch–Kincaid reading ease score above 60 (right of the red line). The overall Flesch–Kincaid reading ease scores corresponded to a Flesch–Kincaid grading level, with a bimodal distribution of 27 web sites (36%) and 29 web sites (39%) with a grade level of 8th to 10th and 10th to 12th grades, respectively (Fig. 1B). Similarly, the Gunning fog index had a bimodal distribution with 24 (32%) and 31 (41%) web sites with a grade level between 10th to 12th and 12th to 14th grades, respectively (Fig. 1C). Only 9 web sites (12%) were below the eighth grade reading level when using the SMOG index (Fig. 1D). The Coleman–Liau index had the highest grade level distribution with 49 (65%) of the 75 web sites with a reading grade level of 12th to 14th grade (Fig. 1E). The automated readability index showed similar results to the Gunning fog score; however, 6 web sites (8%) were below the eighth grade level (Fig. 1F).

Web sites were then categorized into institutional or private web sites. Of the 75 web sites, only 24 web sites (32%) were from institutional sources, while the remaining 51 (68%) were from private sources. When comparing the DISCERN, HONcode, and JAMA scores between institutional and private web sites (Table 2), institutional web sites had a higher DISCERN score (3.4 ± 0.1 vs. 3.1 ± 0.1; \( P = 0.039 \)) and HONcode score (10.3 ± 0.5 vs. 8.8 ± 0.4; \( P = 0.032 \)). However, no statistical difference in the JAMA score (2.1 ± 0.2 vs. 1.8 ± 0.1; \( P = 0.26 \)) was observed. Technical quality also deviated depending on the source, with institutional web sites scoring 0.85 ± 0.02 and private web sites scoring 0.78 ± 0.01 (\( P = 0.004 \)). The differences in technical quality between institutional and private web sites were further analyzed by evaluating each factor (Fig. 2). A higher proportion of institutional web sites had “No Ads in the Text” (100% vs. 78.4% ± 4.9%) and “Loaded in 3 seconds” (100% vs. 86.3% ± 4.5%) when compared with private web sites (\( P < 0.05 \)).

Readability scores for all 6 measures were compared between institutional and private web sites (Table 3). Of the 6 measures, institutional web sites had a lower grade level for the Gunning fog score (11.8 ± 0.4 vs. 13.1 ± 0.2; \( P = 0.02 \)) and the SMOG index (9.0 ± 0.4 vs. 9.8 ± 0.2; \( P = 0.03 \)). However, neither web site source had an average grade level of eighth grade or below nor a Flesch–Kincaid reading level greater than 60.

**DISCUSSION**

Patients are using the internet more frequently to find information about their health conditions because of the abundance of information, the convenience of obtaining it, and the anonymity they can maintain throughout the process.\(^3^0\) Despite these attributes, information on the internet must be approached with caution. It is important that the information being consumed by patients is of the highest quality, reliable, requiring no technical difficulties, and written in language that can be easily understood by most of the adults. In the current investigation, the results show that web sites on DED maintain an adequate quality and reliability of information, poor readability, and good technical quality and therefore require overall improvement.

Our results found an average DISCERN score (±SE) of 3.2 ± 0.1, indicating that the sites have “potentially important but not serious shortcomings.”\(^3^1\) DISCERN relays the quality of health information presented by judging it on various factors such as the sources of information clearly listed, being balanced and unbiased, are every treatment option thoroughly explained with both benefits and risks listed, and does the site support shared decision-making between the patient and their physician. The ideal web site would score a perfect 5, yet none of the sites analyzed achieved that feat, even those developed by reputable academic institutions and organizations.

The HONCode measures the reliability and credibility of information by asking questions such as were the qualifications of the author indicated and does the information support the doctor–patient relationship. By scoring 2 for each criterion of the HONcode (indicating they fully meet that principle) and reaching a max total score of 16, web sites can prove themselves as an honest, credible source of information. In the current investigation, there was a wide range of scores with an average of 9.3 ± 0.3. Some sites reached a max score of 16 and have even received certification from the Health on the Net foundation as in compliance with the HONCode.\(^3^2\) However, other sites had very low scores of 5 or 6, without sharing any information on the qualifications of the authors or citing the sources used. Web sites that have the HONCode badge of certification should be sought out by patients and physicians to obtain information that has been properly vetted. However, not all web sites undergo a HONcode evaluation, which puts the burden on patients and physicians to evaluate these web sites themselves. Although many patients and physicians are capable of doing so, formal evaluation of health care web sites should be performed by the Health on the Net foundation. Furthermore, web site managers of health care web sites should seek out approval by HONcode to increase the credibility of their web sites and

---

**TABLE 1. Quality & Reliability Analysis Compared Between Independent Reviewers**

| Criterion (Range) | Reviewer 1 (±SE) | Reviewer 2 (±SE) | \( P \) |
|------------------|-----------------|-----------------|------|
| DISCERN (1–5)    | 3.1 ± 0.1       | 3.3 ± 0.1       | 0.19 |
| HONcode (0–16)   | 9.2 ± 0.3       | 9.4 ± 0.4       | 0.79 |
| JAMA (0–4)       | 2.0 ± 0.1       | 1.9 ± 0.1       | 0.60 |
| Technical quality (0–1) | 0.81 ± 0.01 | 0.79 ± 0.01 | 0.25 |
ensure patients that they are receiving high quality and reliable information.

The 4 main tenets of the JAMA criteria include authorship, attribution, disclosure, and currency. Our results found a wide range of scores with some scoring the max score of 4, meeting every tenet, and some scoring 0, not meeting any. However, the mean JAMA score (±SE) was 1.9 ± 0.1 indicating that half of the tenets are not being met. This raises red flags when considering the credibility of most sites. The 2 tenets not being met most frequently include authorship, that is, clearly stating the authors and relevant credentials, and attribution, that is, citing all references and sources used for the web site’s content. 66% and 77% of all the web sites evaluated did not meet the criteria of authorship or attribution, respectively, which is exceedingly low for 2 criteria that both have very simple solutions. Therefore, many of these web sites can improve their overall quality and reliability by just adding information of the sources used and the author of the web site’s content.

Most sites scored well when it related to technical qualities (0.8 ± 0.02) which were determined by evaluating the presence of 10 features as previously described.14 This indicates that it is likely easy for patients to access and navigate each web site. One of the shortcomings for most sites was not being up to date (updated within a year) and lacking access to a blog. Institutional sites also scored significantly better than private sites overall and specifically in the categories of “not having ads within the text” and “able to load the webpage within 3 seconds” (Fig. 2). Nonetheless,

TABLE 2. Analysis of Quality, Reliability, and Technical Quality Between Institutional and Private Web Sites

| Criterion (Range)             | Institutional | Private | P   |
|------------------------------|---------------|---------|-----|
| DISCERN (1–5)                | 3.4 ± 0.1     | 3.1 ± 0.1 | 0.039 |
| HONcode (0–16)               | 10.3 ± 0.5    | 8.8 ± 0.4 | 0.032 |
| JAMA (0–4)                   | 2.1 ± 0.2     | 1.8 ± 0.1 | 0.26  |
| Technical quality (0–1)      | 0.85 ± 0.02   | 0.78 ± 0.01 | 0.004 |

*P < 0.05 versus private web sites.
both institutional and private web sites had very low proportions of web sites with up-to-date material. This is crucial not only for patients that need the most up-to-date material when learning about a certain health condition but also for medical professionals and researchers that require up-to-date material to provide the highest quality information.33 The importance of up-to-date medical information has never been more apparent, especially with the onset of the COVID-19 pandemic, in which fast, high quality, reliable information was and still is a desired commodity.34,35

Another important measure in the assessment of online health information is the readability of the text provided. As stated earlier, the average reading level for an American adult is between the seventh and eighth reading grade level.10 The metrics used to assess readability in this study include the Flesch–Kincaid reading ease score & grade level, the Gunning fog score, the SMOG index, the Coleman–Liau index, and the automated readability index, all of which have been used to assess the readability of online health information.26 However, the overwhelming majority of web sites, and with some metrics, 100% of the web sites, had an average reading grade level above the eighth grade (Fig. 1). Furthermore, 76% of all web sites had a Flesch–Kincaid reading ease score below 60, which is classified as “difficult” by the US Department of Health and Human Services.37 Although this does not preclude a reader of lower reading level from using these web sites, it lowers comprehension, which can affect decision-making and patient expectations.38 When comparing institutional and private sites, institutional web sites managed to maintain a lower reading level than private sites for the Gunning fog score ($P = 0.02$) and SMOG index ($P = 0.03$). However, even with this improved readability, the institutional web sites still had a higher than eighth grade reading level for every single metric used (Table 3).

The present investigation provides unique insight into the status of online information on DED. However, there are a couple of limitations that need to be addressed. One minor limitation is the use of Google as the primary search engine. As of February 2021, Google holds 87% of the worldwide market search engine market share,39 but the presence of Bing and Yahoo and other similar search engines provide different avenues for information and would therefore require similar evaluation. Another limitation is the subjective nature of the DISCERN and HONcode metrics. Grading web sites using these criteria require the grader to make a judgment about how well the web site satisfies the respective criteria. However, because there was no statistically significant difference between the 2 independent graders in this study (Table 1), then it is unlikely that the subjective nature of the grading led to any variation.

DED has a major impact on the quality of life of patients19 and a huge economic burden with an estimated annual cost of $55 billion to the US economy.21 Providing patients with reliable, quality data that can be easily understood is paramount in helping to mitigate their disease and improve their quality of life. However, quality, reliability, and readability scores were low for most web sites. Although institutional web sites achieved higher scores than private web sites, revision is warranted for all web sites to improve their overall quality of information and readability profile for the benefit of the patient.

### TABLE 3. Readability Assessment of Institutional and Private Web Sites

| Readability Metric (Range) | Institutional | Private | $P$  |
|---------------------------|--------------|---------|------|
| Flesch–Kincaid ease (0–100) | 57.8 ± 2.8 | 53.3 ± 1.0 | 0.13 |
| Flesch–Kincaid grade level (0–18) | 9.6 ± 0.5 | 10.6 ± 0.2 | 0.05 |
| Gunning fog score (0–18) | 11.8 ± 0.4* | 13.1 ± 0.2 | 0.02 |
| SMOG index (0–18) | 9.0 ± 0.4* | 9.8 ± 0.2 | 0.004 |
| Coleman–Liau index (0–18) | 12.9 ± 0.5 | 13.0 ± 0.1 | 0.80 |
| Automated readability index (1–14) | 10.2 ± 0.6 | 11.3 ± 0.2 | 0.10 |

* $P < 0.05$ versus private web sites.
