The Quality of Mental Health Information Commonly Searched For on the Internet

John M. Grohol, PsyD1; Joseph Slimowicz, MA2; and Rebecca Granda, MSEd3

Abstract

Previous research has reviewed the quality of online information related to specific mental disorders. Yet, no comprehensive study has been conducted on the overall quality of mental health information searched for online. This study examined the first 20 search results of two popular search engines—Google and Bing—for 11 common mental health terms. They were analyzed using the DISCERN instrument, an adaptation of the Depression Website Content Checklist (ADWCC), Flesch Reading Ease and Flesch–Kincaid Grade Level readability measures, HONCode badge display, and commercial status, resulting in an analysis of 440 web pages.

Quality of Web site results varied based on type of disorder examined, with higher quality Web sites found for schizophrenia, bipolar disorder, and dysthymia, and lower quality ratings for phobia, anxiety, and panic disorder Web sites. Of the total Web sites analyzed, 67.5% had good or better quality content. Nearly one-third of the search results produced Web sites from three entities: WebMD, Wikipedia, and the Mayo Clinic. The mean Flesch Reading Ease score was 41.21, and the mean Flesch–Kincaid Grade Level score was 11.68. The presence of the HONCode badge and noncommercial status was found to have a small correlation with Web site quality, and Web sites displaying the HONCode badge and commercial sites had lower readability scores. Popular search engines appear to offer generally reliable results pointing to mostly good or better quality mental health Web sites. However, additional work is needed to make these sites more readable.

Introduction

Over the past 15 years, the number of studies examining people’s health seeking behaviors has doubled during each of the three 5-year periods. Eight out of every 10 Internet users has searched for health information online—approximately 59% of the entire U.S. population and more than 74% of college students. Sixty-nine percent of Swiss psychiatric outpatients and 35% of U.S. adults reported using the Internet to search for a medical or health condition for themselves or someone they know.

Online health information initially garnered a bad reputation for its variable and often poor quality. Some researchers have proposed specific criteria to help rate the quality of a Web site’s information. Some of the more popular rating instruments include DISCERN, Silberg’s accountability criteria: authorship, attribution, disclosure, and currency; Eysenbach’s criteria: accuracy, completeness, readability, design, and information presentation; the Web-MedQual Scale; and Godin’s Quality Assurance Rating Tool for Internet Health Sites (Version 3). In two studies, Jada and Gagliardi reviewed 47 and 98 different instruments used to rate health information online. Only a few provided interrater reliability and construct validity data; most have not been used again.

The most widely used review system has, therefore, become the DISCERN instrument. It was developed in the United Kingdom through a process involving an expert panel, health information providers, and patients from self-help groups. Although not designed specifically for online health information, its simplicity and short length has lent to its widespread use in research focused on judging the quality of medical and mental condition Web sites. It is a reliable instrument with good interrater reliability. Previous research has found the DISCERN to be a reliable indicator of evidence-based Web site quality.

No single study has previously examined the quality of health information online for the most commonly diagnosed and searched for mental disorders in the United States. Nor has any study examined interrater reliability in their assessment of the quality of such Web sites. It has been more than 4 years since a comprehensive review was conducted about...
depression information on the Internet, more than 3 years for bipolar disorder resources, and more than 2 years for schizophrenia resources. Furthermore, the comparison of results between studies such as these is complicated, as researchers use a variety of rating methods and instruments to assess quality, and researchers use widely variable methods to conduct Internet searches to identify the Web sites to evaluate.

Too often, health information is written at a grade level that far exceeds the standard level of readability in the population. McInnes and Haglund found in their review of health Web sites that a single Web site was “written at or below the recommended sixth grade reading level, with ‘easy’ or ‘very easy’ readability,” and that “45 percent of all sites were ‘difficult’ to read.” They also noted that a common online health resource—Wikipedia—had one of the poorest readability scores, putting it close to the “very difficult” to read cutoff. Other researchers have come to similar conclusions, namely that readability of health Web sites is an ongoing concern.

As the Internet gets more crowded, competing for users’ limited attention on mobile, tablet, and other computing platforms, companies have blurred the lines between editorial and sponsored content, further confusing and complicating the search for reliable, quality health information online. With the dizzying array of 273 badges, accreditation organizations, trustmarks, seal of approvals, and other methods for determining quality of health information online, continuously updated research is needed to evaluate the quality of health information online.

The specific aims of the study were to (a) identify popular mental health resources using the two most commonly used search engines; (b) review the quality of these resources using predefined criteria; and (c) assess their readability.

Materials and Methods

This study sought to replicate and update the findings of Ferreira-Lay and Miller, Barnes et al., and Lissman and Boehnlein. However, rather than conduct a single condition-specific study, the present study decided to examine the quality and readability of Web sites returned for 11 of the most commonly diagnosed and searched for mental disorders in the United States.

Most users (77%) start their search for health information online with a search engine, such as Google, Bing, or Yahoo. We therefore conducted Internet keyword searches on the two most popular search engines: Google and Bing. According to comScore, an online metrics company, Google constitutes approximately 67% of the search market in the United States, while Bing and Bing-powered sites (such as Yahoo) constitute approximately 29%. These two search engines therefore constitute the vast majority (>95%) of the U.S. Internet search market.

Measures

DISCERN. The DISCERN instrument is a reliable questionnaire developed by a team of experts to standardize and rate quality information—with the need for content expertise—for consumer health treatment choices. Previous studies suggest DISCERN is a reliable indicator of health information quality. Professional and consumer ratings using the DISCERN have been found to be significantly correlated. The DISCERN instrument consists of 16 questions that are rated on a 5-point Likert scale, with higher scores indicating higher quality of health information.

Adapted Depression Website Content Checklist (ADWCC). The Depression Website Content Checklist (DWCC) was created as a simple tool to evaluate quality of depression information found on Web sites and has good psychometric properties. The DWCC has 10 checklist items covering symptoms, prognosis, and treatment that are scored as either present or absent for the information being rated, resulting in a total score ranging from 0 to 10. This checklist was adapted for the present study based on the 11 disorders evaluated, with the symptoms drawn from the Diagnostic and Statistical Manual of Mental Disorders (4th ed.). The adapted checklist included whether specific symptoms are present on the Web site, psychotherapy and medications are discussed, side effects are noted, and referral information provided.

Readability. Readability was calculated by Microsoft Word 2010 resulting in two variables: the Flesch–Kinsaid readability ease score, and grade level. Higher readability ease scores and lower grade level scores indicate content that is easier to read.

Commercial status. The raters of the study evaluated whether the Web site being rated was either of commercial or noncommercial status. Commercial Web sites were defined as business oriented, having advertisements, and/or for profit. Noncommercial Web sites were defined as nonprofit or government Web sites.

HONCode. Health on the Net Foundation established voluntary guidelines for “medical information you can trust.” The code of ethical content spans eight principles: the information is authoritative, complementary to the doctor–patient relationship, confidentiality is supported, attribution is cited, claims are justified, it is transparent, financial disclosure is made, and advertising is clearly marked. If a site adheres to these principles, they may voluntarily display the HONCode badge.

Selection of Web sites. To simulate an Internet search that was likely to be undertaken by clinicians or patients, 11 of the most prevalent mental disorders were examined in both tools, and the most commonly used term was chosen. Common words that are not exclusive to mental health were not used (such as “ADD” or “sad”). For example, for attention deficit hyperactivity disorder, the following terms were tested: ADHD, attention disorder, attention deficit disorder, attention deficit hyperactivity disorder, attention hyperactivity disorder, and attention hyperactivity. Based upon this analysis, the following 11 search terms were selected...
for research: depression, anxiety, ADHD, bipolar, schizophrenia, dysthymia, phobia, social anxiety, post-traumatic stress disorder (PTSD), panic, and generalized anxiety disorder.

The search was conducted over a 3 month time period from May through July 2012 on the Google and Bing search engines. The first 20 sites generated by each search engine from an anonymized, non-logged-in account were reviewed for each search term, for a total of 440 Web sites. We examined only the first two pages of search results (20 sites) because more than 77% of users only look at the first page of search results, with most remaining users not going beyond the second page of search results.6,35

Only organic search results were examined. Paid placement search results, results manually placed on the page by Google or Microsoft, images, and news items were excluded.

Raters and procedure
Two doctoral graduate students who had previous research and data collection experience were selected as raters. Interrater reliability was established during a pilot study conducted on the search term “depression.” Differences in rating styles and rating conflicts were discussed and resolved. Overall interrater reliability exhibited good interclass correlation (ICC) utilizing a two-way mixed averaged measures (3, K) interclass correlation models with an overall average ICC of 0.73. The pilot spreadsheet was averaged amongst the two raters and included in the final analysis.

After the pilot study was completed and sufficient reliability was established, each rater worked independently rating half of the remaining disorder keyword results based upon the DISCERN, ADWCC criteria, readability, HONCode badge, and commercial status.

Exploratory analyses and descriptive analyses were conducted to determine correlational relationships among the variables in this study. The method of search engine ranking, commercial usage, reading ease and grade level, and quality of information were the variables initially examined to evaluate whether quality of mental health information found online varies based on these determinants.

Results
The statistical software SPSS v20 (IBM Corp., Armonk, NY) was used for the data analyses. Descriptive statistics indicated that Web sites scored highest on the DISCERN instrument on the following questions: clarity of the availability of more than one treatment option, publication achieving its aims, and providing additional sources of support. The lowest three scored questions were: risks of each treatment, reference to uncertainty, and description of effects of no treatment. Means and standard deviations for all questions are reported in Table 1.

Analyses were run to evaluate differences in scores across diagnostic categories. Descriptive statistics indicated that overall DISCERN scores were highest for the diagnostic categories of dysthymia, bipolar disorder, and schizophrenia. The lowest scoring diagnostic categories on the total scores were for phobia, anxiety, and panic disorder. Means and standard deviations for all disorders’ scores are reported in Table 2.

Overall mean scores on the ADWCC for each disorder were compared. Descriptive statistics indicated that Web sites offering information about PTSD, schizophrenia, and depression included the highest amount of symptom information. Disorders pertaining to information about phobias, anxiety, and ADHD contained the least amount of symptom information. Means and standard deviations for all disorders’ ADWCC scores are reported in Table 3.

The level of convergence between the DISCERN and ADWCC instruments was examined, indicating a strong correlation between the two measures’ total scores, \( r(338) = 0.79, p < 0.01 \). Analyses of each disorder’s DISCERN and ADWCC scores demonstrated strong correlations (ranging from 0.78 to 0.89), which suggest strong convergent validity between these two measures.

Of the 440 Web sites examined, 51.1% of them were classified as commercial, 33.2% as nonprofit, 14.8% as governmental, with the remaining 0.9% overseen by a pharmaceutical company (see Fig. 1). The majority of Web sites analyzed (71.4%) did not display the HONCode.

To examine the relationship between the presence of the HONCode badge and Web site quality, a point biserial correlation was conducted. The presence of a HONCode badge was typically associated with higher quality scores. If a HONCode badge was present, the mean scores were higher across all categories.

### Table 1. Means and Standard Deviations of DISCERN Instrument Categories

<table>
<thead>
<tr>
<th>Category</th>
<th>n</th>
<th>Mean</th>
<th>Standard deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>More than one treatment</td>
<td>440</td>
<td>3.82</td>
<td>1.67</td>
</tr>
<tr>
<td>Aims achieved</td>
<td>398*</td>
<td>3.74</td>
<td>1.25</td>
</tr>
<tr>
<td>Details of support</td>
<td>440</td>
<td>3.74</td>
<td>1.69</td>
</tr>
<tr>
<td>Balanced and unbiased</td>
<td>440</td>
<td>3.68</td>
<td>1.38</td>
</tr>
<tr>
<td>Relevance</td>
<td>440</td>
<td>3.43</td>
<td>1.28</td>
</tr>
<tr>
<td>Clear when published</td>
<td>440</td>
<td>3.35</td>
<td>1.72</td>
</tr>
<tr>
<td>Clear aims</td>
<td>440</td>
<td>3.33</td>
<td>1.17</td>
</tr>
<tr>
<td>Describes treatment</td>
<td>440</td>
<td>3.20</td>
<td>1.52</td>
</tr>
<tr>
<td>Benefits of treatment</td>
<td>440</td>
<td>3.11</td>
<td>1.56</td>
</tr>
<tr>
<td>Clear sources</td>
<td>440</td>
<td>2.83</td>
<td>1.62</td>
</tr>
<tr>
<td>Overall quality of life</td>
<td>440</td>
<td>2.64</td>
<td>1.27</td>
</tr>
<tr>
<td>Decision making support</td>
<td>440</td>
<td>2.35</td>
<td>1.59</td>
</tr>
<tr>
<td>Risks of treatment</td>
<td>440</td>
<td>2.28</td>
<td>1.46</td>
</tr>
<tr>
<td>Refer to uncertainty</td>
<td>440</td>
<td>2.25</td>
<td>1.34</td>
</tr>
<tr>
<td>No treatment explored</td>
<td>440</td>
<td>1.99</td>
<td>1.50</td>
</tr>
<tr>
<td>Valid N (listwise)</td>
<td>398</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*Note: The “Aims achieved” category (n = 398) is due to some Web sites not receiving a rating in this category. The DISCERN instrument instructs that no rating should be assigned if there is an insufficient score assigned to the “clear aims” category.

### Table 2. Means and Standard Deviations of DISCERN Diagnostic Categories

<table>
<thead>
<tr>
<th>Category</th>
<th>n</th>
<th>Mean</th>
<th>Standard deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Phobia</td>
<td>40</td>
<td>37.20</td>
<td>15.64</td>
</tr>
<tr>
<td>Anxiety</td>
<td>40</td>
<td>43.15</td>
<td>20.57</td>
</tr>
<tr>
<td>Panic</td>
<td>40</td>
<td>43.30</td>
<td>17.45</td>
</tr>
<tr>
<td>GAD</td>
<td>40</td>
<td>48.25</td>
<td>13.99</td>
</tr>
<tr>
<td>ADHD</td>
<td>40</td>
<td>48.75</td>
<td>16.86</td>
</tr>
<tr>
<td>Social phobia</td>
<td>40</td>
<td>49.75</td>
<td>14.36</td>
</tr>
<tr>
<td>Depression</td>
<td>40</td>
<td>52.03</td>
<td>15.08</td>
</tr>
<tr>
<td>PTSD</td>
<td>40</td>
<td>51.78</td>
<td>13.16</td>
</tr>
<tr>
<td>Dyshymia</td>
<td>40</td>
<td>52.10</td>
<td>17.65</td>
</tr>
<tr>
<td>Bipolar</td>
<td>40</td>
<td>52.33</td>
<td>17.86</td>
</tr>
<tr>
<td>Schizophrenia</td>
<td>40</td>
<td>52.80</td>
<td>17.67</td>
</tr>
</tbody>
</table>
was significantly related to total DISCERN scores, \( r(338) = 0.32, p < 0.01 \) and total ADWCC scores, \( r(338) = 0.31, p < 0.01 \). To examine the relationship between commercial status and Web site quality, a correlational analysis indicated a negative relationship between commercial status and overall DISCERN scores, \( r(338) = -0.22, p < 0.01 \), and total ADWCC scores, \( r(338) = -0.11, p < 0.01 \).

Readability across all Web sites examined found a mean Flesch Reading Ease score of 41.21 and a mean Flesch–Kincaid Grade Level score of 11.68 (see Fig. 2).

The results indicate that there was a small but significant relationship between HONCode status and grade level, \( r(395) = 0.11, p < 0.05 \), as well as HONCode status and Flesch Reading Ease score, \( r(395) = 0.13, p < 0.05 \).

There was a small yet significant inverse relationship between commercial status and Flesch–Kincaid Reading Ease score, \( r(395) = -0.15, p < 0.01 \). A weak yet significant relationship was also observed between commercial status and Flesch–Kincaid grade level, \( r(395) = 0.10, p < 0.05 \). Results indicated that commercial status was associated with a decreased reading ease score and higher grade level score.

A DISCERN score of 40 or better as a cutoff criteria was suggested by Khazaal et al.\(^{31}\) on the basis of statistical analyses evaluating sensitivity and specificity of the measure. Out of the 440 Web sites analyzed, 297 (67.5%) had DISCERN scores of \( \geq 40 \). Web sites that had two or more Web site pages in the top five totals across all mental disorders included: Help Guide.org, Mayo Clinic, National Institute of Mental Health, Psych Central, Wikipedia, and the WebMD sites eMedicineHealth, MedicineNet, and WebMD.

Sixty-eight search results (15.5%) were from a single commercial entity: WebMD or a Web site overseen by WebMD. Thirty-eight Web site pages (8.6%) were from Wikipedia, while 31 Web sites (7.0%) were from the Mayo Clinic. Five Web sites constituted nearly 40% of the search results (see Fig. 3), which included the top two mental health focused Web sites. The commercial site—PsychCentral.com—had 19 pages (4.3%), while the nonprofit site, HelpGuide.org, had 18 pages (4.1%). Seventeen Web sites (3.9%) were from an online dictionary, providing only a minimal definition of the term being researched.

Ownership of most Web sites was obvious in 93.6% of those analyzed, and most had a privacy policy (89.1%). Only 39.3% of Web sites reviewed had an editorial board.

### Discussion

The results of the present study confirm previous findings about the quality of mental health resources online. Search engines regularly returned Web sites that were of good or better quality health information. Following Khazaal et al.’s recommendation,\(^{31}\) a DISCERN score cutoff of \( > 40 \) identified 67.5% of the Web sites examined as having good or better quality content. This finding is also in keeping with Reaveley and Jorm’s\(^ {32} \) conclusion that the quality of mental health information is likely improving over time.

Not all mental health information is represented equally online, however. Some disorders—dysthymia, bipolar disorder, and schizophrenia—had higher quality information presented than other disorders—phobia, anxiety, and panic disorder. However, except for phobias, all disorders scored a mean of \( \geq 40 \) on the DISCERN instrument. Phobia quality scores may suffer, as they have become part of mainstream culture and popular psychology, often showing up as lists in “phobia” search results.

An examination of specific items on the DISCERN demonstrates that some items are not often discussed on the mental health Web sites we examined: risks of treatment, uncertainty about treatment choices, and the risks and benefits of not getting treatment. This could reflect a pro-treatment bias on mental health Web sites, which in turn may be a result of Web site authors attempting to counter stigmatization and undertreatment of mental disorders. For instance, 60% of those with a mental disorder over a 12 month period failed to receive treatment.\(^ {36} \)

Our findings on the use of Web site characteristics—such as the HONCode or commercial status—as a quality indicator contribute to the mixed literature on these signs. The present study found a weak but positive relationship between health information quality and the display of the HONCode. However, other studies have found little support for using the HONCode as a quality indicator.\(^ {31,37} \) A small correlation was also uncovered suggesting that noncommercial Web sites tended to score higher on the DISCERN and ADWCC measures. This is consistent with prior research findings that nonprofit organizational Web sites tended to score higher on content quality.\(^ {26} \) However, others have found no relationship between quality of health information and commercial

### Table 3. Means and Standard Deviations of ADWCC Diagnostic Categories

<table>
<thead>
<tr>
<th>Category</th>
<th>n</th>
<th>Mean</th>
<th>Standard deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Specific phobia</td>
<td>40</td>
<td>3.03</td>
<td>3.38</td>
</tr>
<tr>
<td>Anxiety</td>
<td>40</td>
<td>5.10</td>
<td>3.85</td>
</tr>
<tr>
<td>ADHD</td>
<td>40</td>
<td>5.78</td>
<td>3.74</td>
</tr>
<tr>
<td>Dysthymia</td>
<td>40</td>
<td>6.15</td>
<td>2.81</td>
</tr>
<tr>
<td>Panic</td>
<td>40</td>
<td>6.35</td>
<td>3.13</td>
</tr>
<tr>
<td>Social anxiety</td>
<td>40</td>
<td>6.85</td>
<td>2.42</td>
</tr>
<tr>
<td>GAD</td>
<td>40</td>
<td>6.88</td>
<td>2.24</td>
</tr>
<tr>
<td>Bipolar</td>
<td>40</td>
<td>7.43</td>
<td>3.07</td>
</tr>
<tr>
<td>Depression</td>
<td>40</td>
<td>7.43</td>
<td>2.06</td>
</tr>
<tr>
<td>Schizophrenia</td>
<td>40</td>
<td>7.49</td>
<td>2.65</td>
</tr>
<tr>
<td>PTSD</td>
<td>40</td>
<td>7.90</td>
<td>2.84</td>
</tr>
</tbody>
</table>

ADWCC, Adapted Depression Website Content Checklist.
Given the mixed research findings on these Web site characteristics, it remains unlikely they alone provide a person with a reliable indicator of a Web site’s content quality.

Significantly, readability remains an ongoing concern among mental health Web sites. Results indicate that the presence of a HONCode badge is correlated with a small but significant decrease in reading ease and a similar increase in reading grade level. With an average Flesch Reading Ease score squarely in the “difficult” range of readability and an 11th grade reading level score, most Web sites appear to still struggle with producing mental health content at an 8th grade reading level. This echoes other researchers’ findings, suggesting publishers of mental health information would benefit from focusing future efforts on making information more readable at a lower grade level.23,31

The present study has several limitations. Because of the fluid and ever-changing nature of the Internet and search engines’ algorithms, the results obtained in this study may be impacted over time. The study does not reflect the variability found with regard to individual search engine choices, methods, and phrases chosen. Although both the DISCERN and ADWCC instruments were designed with inexperienced raters in mind, the findings may nonetheless reflect different scores than if experienced raters were used. Since the readability measures employed do not actually measure reading comprehension, they can only act as a proxy for actual comprehension of mental health material.

This study was designed to assess the quality of mental health Web sites found in common search results using a variety of measures and methods. The overall DISCERN and ADWCC scores demonstrate that most of the mental health terms Internet users commonly search for produce generally good or better quality results. The present study confirms that the use of a single Web site characteristic—such as the display of the HONcode badge or commercial status—is unlikely to be a reliable indicator alone of Web site quality, and that readability remains a challenging issue for mental health Web sites. These findings taken together as a whole suggest that search engines’ algorithms for returning relevant, good quality search results can largely be trusted in this specific health content domain.

Author Disclosure Statement

John M. Grohol is the owner and publisher of a mental health Web site, Psych Central, some of whose resources were independently rated in this study. Joseph Slimowicz and Rebecca Granda report no conflicts of interest.

References

2. Fox S. (2011) Health topics: 80% of Internet users look for health information online. Pew Internet & American Life...

Address correspondence to:
Dr. John M. Grohol
Psych Central
55 Pleasant Street
Newburyport, MA 01950

E-mail: grohol@psychcentral.com