Research Paper

Indicators of Accuracy of Consumer Health Information on the Internet:
A Study of Indicators Relating to Information for Managing Fever in Children in the Home

DON FALLIS, PHD, MARTIN FRICKÉ, PHD

Abstract
Objectives: To identify indicators of accuracy for consumer health information on the Internet. The results will help lay people distinguish accurate from inaccurate health information on the Internet.

Design: Several popular search engines (Yahoo, AltaVista, and Google) were used to find Web pages on the treatment of fever in children. The accuracy and completeness of these Web pages was determined by comparing their content with that of an instrument developed from authoritative sources on treating fever in children. The presence on these Web pages of a number of proposed indicators of accuracy, taken from published guidelines for evaluating the quality of health information on the Internet, was noted.

Main Outcome Measures: Correlation between the accuracy of Web pages on treating fever in children and the presence of proposed indicators of accuracy on these pages. Likelihood ratios for the presence (and absence) of these proposed indicators.

Results: One hundred Web pages were identified and characterized as “more accurate” or “less accurate.” Three indicators correlated with accuracy: displaying the HONcode logo, having an organization domain, and displaying a copyright. Many proposed indicators taken from published guidelines did not correlate with accuracy (e.g., the author being identified and the author having medical credentials) or inaccuracy (e.g., lack of currency and advertising).

Conclusions: This method provides a systematic way of identifying indicators that are correlated with the accuracy (or inaccuracy) of health information on the Internet. Three such indicators have been identified in this study. Identifying such indicators and informing the providers and consumers of health information about them would be valuable for public health care.


Millions of people now use the Internet to gather health information. It is important for the health and well-being of these people, and for the societies to which they belong, that the information they acquire in this way be, on the whole, accurate. Much of the consumer health information on the Internet is accurate. However, studies have shown that some of this information is not accurate. The present study is part of a project to help lay people distinguish accurate from inaccurate consumer health information on the Internet.

In a recent study, Impicciatore et al. looked at the accuracy of information on the treatment of fever in children. According to them, “only a few of the [41] Web pages we reviewed gave complete and accurate information for such a common and widely discussed condition.” Thus, Impicciatore et al. established that there is a problem with inaccurate consumer health information on the Internet.
To help Internet users to avoid this inaccurate information, a number of authors and organizations have published guidelines for evaluating the quality of health information on the Internet. These guidelines typically include lists of indicators that are intended to help Internet users determine the accuracy of Web sites. For example, the medical credentials of the author of a Web site are supposed to be an indicator of accuracy. Also, an out-of-date Web site and a lack of advertising on a Web site are supposed to be indicators of inaccuracy.

Of course, these guidelines can only help Internet users to avoid inaccurate information if the indicators really are correlated with accuracy (or inaccuracy). Unfortunately, there is currently no empirical data to support the claim that these indicators are correlated with accuracy (or inaccuracy). In fact, a recent study casts doubt on the reliability of at least one of the proposed indicators of accuracy (namely, positive ratings from services that evaluate medical information on the Internet).

The aim of the current study is to test empirically several of the proposed indicators of accuracy. Following Impicciatore et al., we looked at Web sites that discuss the treatment of fever in children. However, in addition to assessing the accuracy of these sites, we looked for the presence of several of the proposed indicators. With this data (which were collected from March to November 2000), we were able to determine that some of the proposed indicators are correlated with accuracy, but many are not.

**Design**

**Search Protocol**

Web sites on the treatment of fever in children were collected. These sites were found through keyword searches on the following Internet search engines: Yahoo, AltaVista, and Google. In particular, the terms “fever,” “treatment,” and “child” (connected by the Boolean operator “and”) were used. The goal was to search for health information on the Internet in a manner that might be used by a lay person who needed information on treating a child with a fever. An attempt was made to find as many Web sites on this topic as possible.

Sites were selected that specifically undertook to provide recommendations for the treatment of fever in children. Sites that concerned the treatment of fever associated with specific named diseases or conditions were excluded. Also, sites that simply warned against using aspirin to treat a child with a fever were excluded (for such a warning alone does not amount to a recommended course of treatment).

**Accuracy Protocol**

To determine the accuracy of the Web sites, the information on the Web sites was compared with the recommendations of authoritative sources on the treatment of fever in children.

The treatment of fever in children was chosen for this study specifically because there is wide consensus among experts in this area. (The only issues of some debate are the optimal sites for measuring temperature and whether to treat mild fever.) It is generally considered legitimate to judge accuracy by consulting authoritative sources when there is such a consensus among experts. This is, for example, the procedure used by Impicciatore et al.

An instrument was developed that consisted of 25 questions covering the following five topics:

- The minimum temperature considered to indicate fever
- The optimal site(s) for measuring temperature
- Pharmacologic treatment of fever
- Physical treatment of fever
- Conditions that warrant consulting a physician

Fever in Pediatric Practice, by El-Radhi and Carroll and the Merck Manual of Medical Information were used to develop this instrument. This instrument was checked by a physician.

Two observers independently applied this instrument to each Web site and recorded the results in a spreadsheet. For each of the 25 questions, a Web site received 1 point for a completely correct answer, 0.5 points for no answer, and 0 points for an incorrect answer. After the results were recorded, they were double-checked to eliminate typing and transcription errors. Interobserver reliability was assessed by calculating proportions of specific agreement. The proportions of specific agreement test is superior to the more commonly seen kappa test for the reasons given by Cicchetti and Feinstein.

An overall accuracy score (between 0 and 1) was computed for each Web site by taking a weighted average of the scores on the individual questions. (Since more information was involved, the third and fourth topics were weighted more heavily. In addition, Web sites were penalized for including any additional incorrect information on the treatment of fever in children.) This overall accuracy score takes
into account both the correctness and the completeness of the information.

This particular technique for assigning an overall accuracy score to a Web site is based on theoretical work on verisimilitude (or likeness to truth). The verisimilitude of a piece of information (e.g., a scientific theory or the information on a Web site) is its distance from the whole truth on a particular topic. This distance can be calculated in the following manner: The whole truth on the topic is divided into a set of simple, independent, component propositions. The correctness and completeness of the piece of information with respect to each of these component propositions is determined. Finally, the distance of the piece of information from the whole truth is a weighted average of its distance from each of the component propositions.

**Indicators of Accuracy Protocol**

Several proposed indicators of accuracy were taken from published guidelines for evaluating the quality of health information on the Internet. For each Web site, we determined whether each proposed indicator was present or absent. (Interobserver reliability was again assessed here by calculating proportions of specific agreement.) These indicators can be grouped into the following categories:

- Whether the Web site had a commercial domain (e.g., drkoop.com), an organization domain (e.g., kidshealth.org), or an education domain (e.g., columbia.edu)
- Whether the Web site was up to date and whether the treatment page was up to date
- Whether the Web site displayed the HONcode logo (see below)
- Whether the Web site carried any advertising
- Whether the author was identified on the treatment page (and, if so, whether the author was identified as a physician)
- Whether copyright was claimed or acknowledged
- Whether contact information was given
- Whether spelling errors appeared on the treatment page (and, if so, how many)
- Whether there were exclamation points on the treatment page (and, if so, how many)
- Whether peer-reviewed medical literature was cited
- Whether the number of in-links (see below) was high

Two of these proposed indicators of accuracy may require some explanation. First, the HONcode logo is displayed on Web sites that voluntarily agree to abide by the Health On the Net Foundation’s Code of Conduct for publishing quality health information on the Internet. This code of conduct mandates, for example, that “any medical or health advice provided and hosted on this site will only be given by medically trained and qualified professionals” and that such advice will be based on “appropriate, balanced evidence.”

Second, the number of in-links to a particular Web site is the number of other Web sites that include hyperlinks to that Web site. The number of in-links to a Web site is the Web equivalent of the citation count of a journal article from traditional bibliometrics. Several search engines provide information as to how many other sites link to a particular site. In this study, we used Lycos to measure the number of in-links. For the 100 sites that we investigated, the number of in-links varied from 0 to about 37,000. Just less than half of these sites were referred to by more than 1,000 other sites. Web sites to which more than 1,000 other Web sites linked were deemed to have “many in-links.”

**Contingency Table Analysis**

Web sites were divided into two categories on the basis of their overall accuracy scores. Web sites with an overall accuracy score at or above the median were deemed to be the “more accurate” sites. Web sites with an overall accuracy score below the median were deemed to be the “less accurate” sites. Since the interval scores were converted into ordinal scores in this way, nothing hangs on the precise interval value of a Web site’s overall accuracy score.

For each of the proposed indicators, a $2 \times 2$ contingency table was constructed. (For example, Figure 1 is the contingency table for the HONcode logo. Eleven of the 50 “more accurate” sites displayed the HONcode logo, and 39 did not. However, only 3 of the 50 “less accurate” sites displayed the HONcode logo, and 47 did not.) A chi-square test for independence was then used to determine whether the accuracy of Web sites was correlated with the presence (or absence) of the proposed indicators. We concluded that a proposed indicator was correlated with accuracy if the chi-square probability was less than 0.05 (with expected cell frequencies greater than 5).

To ensure that the results of the contingency analysis do not depend on taking the median accuracy score as the dividing line between the “more accurate” Web sites and the “less accurate” Web sites, a sensitivity
analysis was performed. In particular, the same contingency table analysis was performed using several different dividing lines. Also, as an additional check on the results, a point biserial correlation test was performed using the raw accuracy scores. (Since the interval scale for accuracy was a constructed rather than a natural scale, it was not appropriate to use this test as our main statistical tool.)

**Likelihood Ratios**

In addition to determining whether there was a correlation with accuracy, we calculated two likelihood ratios for each proposed indicator. Likelihood ratios are used to measure the definitiveness of a diagnostic test.\(^2\) The likelihood ratio for the presence of an indicator of accuracy is how much more likely it is that the indicator will be displayed on a “more accurate” Web site than on a “less accurate” Web site. This likelihood ratio can be estimated using the relative frequencies of “more accurate” Web sites with the indicator and of “less accurate” Web sites with the indicator. Analogously, the likelihood ratio for the absence of an indicator of accuracy is how much more likely it is that the indicator will not be displayed on a “less accurate” Web site than on a “more accurate” site.

**Results**

One hundred Web sites on the treatment of fever in children were found. The overall accuracy scores for these sites ranged from 0.55 to 0.75. The distribution of these scores is shown in Figure 2. Fifty Web sites had overall accuracy scores on or above the median score of 0.65 and were deemed to be the “more accurate” sites. Fifty Web sites had overall accuracy scores below the median score and were deemed to be the “less accurate” sites.

The contingency table analysis showed three of the proposed indicators to be correlated with accuracy (Table 1). In particular, the contingency tables for the HONcode logo, organization domain, and copyright all yielded a chi-square probability of less than 0.05. For example, the probability that the presence of the HONcode logo is not correlated with accuracy is 0.021 (i.e., about 2 percent). Several other commonly proposed indicators, such as authority (e.g., the author having medical credentials) and currency, were not correlated with accuracy at all.

In the sensitivity analysis, the dividing line between the “more accurate” Web sites and the “less accurate” Web sites was set at 0.01 intervals along the range of overall accuracy scores (i.e., 0.56, 0.57, \ldots, 0.73, 0.74). The same contingency tables analysis was performed using each of these dividing lines. Whenever expected cell frequencies were greater than 5, the HONcode logo, organization domain, and copyright were found to be correlated with accuracy (i.e., the chi-square probabilities were less than 0.05). In addition, the results of the point biserial correlation test corroborated the results of the original contingency table analysis.

Since 22 percent of the “more accurate” Web sites but only 6 percent of the “less accurate” Web sites display the HONcode logo, the likelihood ratio for the presence of the HONcode logo is 3.67. Thus, the presence of the HONcode logo is pretty good evidence of accuracy, because the logo is almost four times more likely to be displayed on a more accurate site than on a less accurate site. In contrast, a failure to claim copyright is evidence of inaccuracy because copyright is two times more likely not to be claimed on a less accurate site than on a more accurate site. Point estimates of both likelihood ratios for the HONCode logo, organization domain, and copyright are shown in Table 2.
For each of the questions on the accuracy protocol, the initial proportions of specific agreement were around 0.90. This increased to 0.95 when typing and transcription errors were removed. For each of the proposed indicators, the proportions of specific agreement exceeded 0.95. In addition, there was complete inter-observer agreement with respect to all the indicators that the present study found to be correlated with accuracy (e.g., the presence of the HONcode logo).

**Discussion**

**Indicators Correlated with Accuracy**

There are indicators that lay people can use to distinguish accurate from inaccurate health information on the Internet. This study has identified three such indicators of accuracy: displaying the HONcode logo, having an organization domain, and displaying a copyright. In particular, the presence of the HONcode logo on a Web site is a fairly good indication that a Web site contains accurate information on the treatment of fever in children. It seems that the quality-control procedures mandated by the Health On the Net Foundation’s Code of Conduct tend to result in the publishing of accurate consumer health information.

### Table 1

The Number of “More Accurate” and “Less Accurate” Web Sites for Each of the Proposed Indicators of Accuracy and the Probabilities of Correlation

<table>
<thead>
<tr>
<th>Proposed Indicator of Accuracy</th>
<th>More Accurate Web Sites</th>
<th>Less Accurate Web Sites</th>
<th>Chi-square Probability</th>
<th>Expected Cell Frequencies</th>
</tr>
</thead>
<tbody>
<tr>
<td>A. Domain</td>
<td>Commercial domain</td>
<td>25</td>
<td>32</td>
<td>0.157</td>
</tr>
<tr>
<td></td>
<td>Organization domain</td>
<td>19</td>
<td>10</td>
<td>0.047</td>
</tr>
<tr>
<td></td>
<td>Education domain</td>
<td>4</td>
<td>4</td>
<td>1.000</td>
</tr>
<tr>
<td>B. Currency</td>
<td>Main page more than 1 yr old</td>
<td>2</td>
<td>2</td>
<td>1.000</td>
</tr>
<tr>
<td></td>
<td>Main page 6 mo–1 yr old</td>
<td>2</td>
<td>3</td>
<td>0.646</td>
</tr>
<tr>
<td></td>
<td>Main page &lt;6 mo old</td>
<td>9</td>
<td>10</td>
<td>0.799</td>
</tr>
<tr>
<td></td>
<td>No date given on main page</td>
<td>37</td>
<td>35</td>
<td>0.656</td>
</tr>
<tr>
<td></td>
<td>Treatment page &gt;1 yr old</td>
<td>5</td>
<td>2</td>
<td>0.240</td>
</tr>
<tr>
<td></td>
<td>Treatment page 6 mo–1 yr old</td>
<td>5</td>
<td>4</td>
<td>0.727</td>
</tr>
<tr>
<td></td>
<td>Treatment page &lt;6 mo old</td>
<td>4</td>
<td>3</td>
<td>0.695</td>
</tr>
<tr>
<td></td>
<td>No date given on treatment page</td>
<td>36</td>
<td>41</td>
<td>0.235</td>
</tr>
<tr>
<td>C. HONcode</td>
<td><em>HONcode logo displayed</em></td>
<td>11</td>
<td>3</td>
<td>0.021</td>
</tr>
<tr>
<td>D. Advertising</td>
<td>Advertising displayed</td>
<td>16</td>
<td>18</td>
<td>0.673</td>
</tr>
<tr>
<td>E. Authorship</td>
<td>Author identified</td>
<td>15</td>
<td>16</td>
<td>0.829</td>
</tr>
<tr>
<td></td>
<td>Author identified as an MD</td>
<td>12</td>
<td>10</td>
<td>0.629</td>
</tr>
<tr>
<td>F. Copyright</td>
<td>Copyright claimed</td>
<td>37</td>
<td>24</td>
<td>0.008</td>
</tr>
<tr>
<td>G. Contact Information</td>
<td>Contact information provided</td>
<td>38</td>
<td>32</td>
<td>0.190</td>
</tr>
<tr>
<td>H. Spelling</td>
<td>More than 2 spelling errors</td>
<td>3</td>
<td>0</td>
<td>0.079</td>
</tr>
<tr>
<td></td>
<td>1 or 2 spelling errors</td>
<td>15</td>
<td>12</td>
<td>0.499</td>
</tr>
<tr>
<td></td>
<td>No spelling errors</td>
<td>32</td>
<td>38</td>
<td>0.190</td>
</tr>
<tr>
<td>I. Exclamation Points</td>
<td>More than 2 exclamation points</td>
<td>1</td>
<td>0</td>
<td>0.315</td>
</tr>
<tr>
<td></td>
<td>1 or 2 exclamation points</td>
<td>8</td>
<td>7</td>
<td>0.779</td>
</tr>
<tr>
<td></td>
<td>No exclamation points</td>
<td>41</td>
<td>43</td>
<td>0.585</td>
</tr>
<tr>
<td>J. References</td>
<td>Peer-reviewed medical literature cited</td>
<td>2</td>
<td>2</td>
<td>1.000</td>
</tr>
<tr>
<td>K. Inlinks</td>
<td>More than 1,000 inlinks to the main page</td>
<td>20</td>
<td>21</td>
<td>0.839</td>
</tr>
<tr>
<td></td>
<td>More than 1,000 inlinks to the treatment page</td>
<td>9</td>
<td>5</td>
<td>0.249</td>
</tr>
</tbody>
</table>

*There were 50 “more accurate” Web sites and 50 “less accurate” Web sites in total. The chi-square probabilities have been rounded to three decimal places. The significance level is 0.05. Those indicators with chi-square probabilities less than 0.05 and expected cell frequencies greater than 5 are in italics.

### Table 2

Likelihood Ratios for the Proposed Indicators of Accuracy that have a High Probability of Correlation

<table>
<thead>
<tr>
<th>Proposed Indicator</th>
<th>Presence of Indicator</th>
<th>Absence of Indicator</th>
</tr>
</thead>
<tbody>
<tr>
<td>Organization domain</td>
<td>1.90</td>
<td>1.29</td>
</tr>
<tr>
<td>HONcode logo displayed</td>
<td>3.67</td>
<td>1.21</td>
</tr>
<tr>
<td>Copyright claimed</td>
<td>1.54</td>
<td>2.00</td>
</tr>
</tbody>
</table>

*NOTE: The likelihood ratios have been rounded to two decimal places.*
While such indicators of accuracy can be extremely valuable tools for evaluating health information on the Internet, it must be noted that the presence of these indicators does not guarantee that a Web site contains accurate information (for the relationship between these indicators and accuracy is a probabilistic one). In other words, the HONCode logo is not a definitive “knowledge hallmark.”\(^{22}\)

In addition, the correlation of the presence of a particular indicator with accuracy does not imply that the absence of that indicator is correlated with inaccuracy. For example, while the presence of the HONCode logo is correlated with accuracy, the absence of the HONCode logo is not correlated with inaccuracy. A Web site that does not display the HONCode logo is only slightly more likely to be “less accurate” than to be “more accurate.” In fact, a large number of sites that have accurate information on the treatment of fever in children do not display the HONCode logo.

Finally, the fact that an indicator is currently correlated with accuracy does not guarantee that it will always remain an indicator of accuracy. In particular, providers of information may be motivated to make sure that their Web sites display such indicators whether their information is accurate or not. As a result of this, such indicators may quickly cease to be indicators of accuracy.\(^{23}\)

What is needed to forestall such a development are indicators that are difficult to “fake.” Unfortunately, it is a very simple matter for anyone to claim copyright for their Web site. Thus, copyright is not a very robust indicator of accuracy. The presence of the HONCode logo, however, is somewhat more difficult, although not impossible, to fake. For example, monitoring procedures are in place that are designed to secure the removal of the HONCode logo from Web sites that fail to comply with the Health On the Net Foundation’s Code of Conduct.\(^{24}\) Thus, the HONCode logo is a much more robust indicator of accuracy.

**Indicators Not Correlated with Accuracy**

This study also indicates that several of the proposed indicators from published guidelines for evaluating the quality of health information on the Internet are not correlated with accuracy (or inaccuracy). For example, a number of published guidelines for evaluating medical information on the Internet suggest that an author having medical credentials is an indicator of accuracy.\(^{10,11}\) However, we found that there was no such correlation.

A number of published guidelines for evaluating medical information on the Internet suggest that citing peer-reviewed medical literature is an indicator of accuracy.\(^{10,11}\) Unfortunately, so few sites in our sample cited peer-reviewed medical literature that we were unable to reliably test this assertion. However, we did find that there was no correlation between the failure to cite peer-reviewed literature and inaccuracy.

No one would expect the mere fact that a Web site is up-to-date to be an indicator of accuracy. However, it might be expected that no date on a Web site or an old date on a Web site would be correlated to some degree with inaccuracy.\(^{11}\) However, we found that there was no such correlation. Of course, this may be partly due to the specific medical topic that we examined. The best information about the treatment of fever in children has remained fairly stable for a number of years. Lack of currency may be correlated with inaccuracy in areas of medicine where the best information is rapidly changing (e.g., AIDS research).

Finally, published guidelines for evaluating medical information on the Internet suggest that the presence of advertising on a Web site is an indicator of inaccuracy.\(^{11}\) However, we found that there was no such correlation. Even so, it might be useful to look specifically at Web sites on which the potential conflict of interest is more clear-cut. For example, there are Web sites providing information about the treatment of fever that are sponsored by companies that sell pharmaceuticals for the treatment of fever. Unfortunately, we did not find enough Web sites of this particular sort to be able draw any conclusions on this issue.

**Future Directions for Research**

The most important way in which this study should be extended is to look at a wider range of health topics. First, this would allow us to confirm that these results are not unique to information on the treatment of fever in children. In addition, looking at a wider range of topics would allow for a larger sample size. The current study suggests that even an exhaustive sample of Web sites on a single topic is often going to be too small to allow us to answer certain important questions. For example, since so few Web sites on the treatment of fever cited peer-reviewed medical literature, we were unable to determine whether such citations are correlated with accuracy.

Also, it would be valuable to look at consumer health information on the Internet in other languages and other countries. In this study, we initially tried to look at Web sites in Spanish on the treatment of fever.
in children. Unfortunately, we were unable to find more than a handful of such sites.

Finally, this study has identified indicators of accuracy that lay people can use to distinguish accurate from inaccurate health information on the Internet. However, it would be ideal to automate this process as much as possible. For example, Price and Hersh\(^2\) have tried to incorporate several of the proposed indicators of accuracy into a “software tool that can automatically assess the quality of consumer health Webpages.” A test for the presence of the HONcode logo, for example, would be fairly easy to incorporate into such a system.

**Conclusion**

It is important for consumers of health information on the Internet to be able to distinguish accurate from inaccurate information. Reliable indicators of accuracy can help them do this. The current study shows that it is possible to test empirically the reliability of proposed indicators of accuracy. The presence of the HONcode logo, for example, turns out to be a fairly good indicator of accurate information about the treatment of fever in children. The empirical evidence does not, however, generally support the published guidelines for evaluating the quality of health information on the Internet. Several proposed indicators, such as authority and currency, do not appear to be good indicators of accurate health information. To confirm this result, the present study needs to be replicated on a wider range of health topics.

The authors thank Clayton Curtis, MD, PhD, who checked the protocol for assessing the accuracy of the Web sites. Dr. Curtis is the Director of Clinical Informatics for the Department of Veterans Affairs in the New England Healthcare System. Graduate assistants in the School of Information Resources and Library Science at the University of Arizona (Tracy Cook, Virginia Cullen, Michelle Drumm, Alba Fernandez, Robert Frasier, Elizabeth Gouwens, Melinda Hardman, Andrew Kaplan, and Ping Situ) collected the data. Margaret Higgins, Gerrard Liddell, and two anonymous reviewers provided a number of helpful suggestions, and Kay Mathiesen provided editorial assistance.

**References**

1. Davis R, Miller L. Millions scour the Web to find medical information. USA Today. Jul 14, 1999:1A.