Since its 2001 launch, Wikipedia (http://www.wikipedia.org/) has become the most popular general reference site on the Internet, ranking 6th globally based on Internet traffic. As of March 2014, it contained more than 31 million articles in 285 languages. Wikipedia’s prominence has been made possible by its fundamental design as a wiki, or collaborative database, allowing all users the ability to add, delete, and edit information at will. However, it is this very feature that has raised concern in the medical community regarding the reliability of the information it contains.
Despite these concerns, Wikipedia has become a popular source of health care information, with 47% to 70% of physicians and medical students admitting to using it as a reference. In actuality, these figures may be higher because some researchers suspect its use is underreported. Although the effect of Wikipedia’s information on medical decision making is unclear, it almost certainly has an influence.

Wikipedia has several mechanisms in place to deal with unverifiable information and vandalism. Because of the frequency of editing and revisions, most instances of vandalism only exist for a few days after being identified, with half of the corrections being posted less than 3 minutes after being identified. One study found that some corrections were made almost instantaneously in 42% of cases. There is a push on Wikipedia to have statements backed by references and unverifiable statements being called out to readers. Haigh observed that, in general, medically related articles on Wikipedia are accompanied by a sufficient amount of reputable citations.

To evaluate Wikipedia’s accuracy, we compared Wikipedia articles on the 10 most costly medical conditions in the United States with recognized peer-reviewed sources.

Methods

The 10 most costly conditions in the United States by public and private expenditure in 2008—the year that the most complete data were available for the present study—were identified from the publicly available database from the Agency for Healthcare Research and Quality. We then identified 10 Wikipedia articles that we believed most closely related to each of those conditions. Because Wikipedia articles are dynamic and subject to frequent changes and updates, we printed the selected articles on April 25, 2012, for our research purposes.
disease, and diabetes mellitus—there was a statistically significant discordance between Wikipedia articles and peer-reviewed sources for dissimilar assertions. The interpretation of the $P$ value is true for similar assertions between the 2 reviewers as well as for dissimilar assertions (Table 3).

Discussion

A few studies\textsuperscript{12,25-27} have compared Wikipedia articles with standard peer-reviewed sources and have shown it to be roughly equivalent to these sources. The most notable study, by Giles,\textsuperscript{25} compared Wikipedia with the Encyclopedia Britannica. Other authors\textsuperscript{12,26,27} have compared Wikipedia with textbooks and national databases and showed comparable results. In contrast, other researchers\textsuperscript{28-30} have determined that Wikipedia is unsuitable as a reference for drugs. Except for psychiatric conditions,\textsuperscript{26} scientific research has never, to our knowledge, focused on Wikipedia’s content on prevalent medical conditions. A recent study by Azer\textsuperscript{31} concluded that Wikipedia is not a reliable information source for medical students in gastroenterology and hepatology.

The present study demonstrated that most Wikipedia articles on the 10 most costly conditions in the United States contained assertions that are inconsistent with peer-reviewed sources. Because our standard was the peer-reviewed published literature, it can be argued that these assertions on Wikipedia represent factual errors.

A perplexing finding in our study was that most of the dissimilar assertions found by the reviewers failed to demonstrate discordance. A reporting bias may have plausibly occurred: each article reviewer was either an internal medicine resident or a rotating intern physician at the time of the review and may not have believed that every assertion was worth reporting. For example, the diabetes mellitus Wikipedia article stated that it is a condition in “which a person has high blood sugar.” One reviewer might have accurately recorded this statement as

Table 1.
Top 10 Most Costly Conditions in the United States\textsuperscript{a} and Corresponding Wikipedia Articles\textsuperscript{b}

<table>
<thead>
<tr>
<th>Conditions</th>
<th>Corresponding Wikipedia Article</th>
</tr>
</thead>
<tbody>
<tr>
<td>Heart disease</td>
<td>Coronary artery disease\textsuperscript{15}</td>
</tr>
<tr>
<td>Cancer</td>
<td>Lung cancer\textsuperscript{16}</td>
</tr>
<tr>
<td>Mental disorders</td>
<td>Major depressive disorder\textsuperscript{17}</td>
</tr>
<tr>
<td>Trauma-related disorders</td>
<td>Concussion\textsuperscript{18}</td>
</tr>
<tr>
<td>Osteoarthritis</td>
<td>Osteoarthritis\textsuperscript{19}</td>
</tr>
<tr>
<td>Chronic obstructive lung disease/asthma</td>
<td>Chronic obstructive pulmonary disease\textsuperscript{20}</td>
</tr>
<tr>
<td>Hypertension</td>
<td>Hypertension\textsuperscript{21}</td>
</tr>
<tr>
<td>Diabetes</td>
<td>Diabetes mellitus\textsuperscript{22}</td>
</tr>
<tr>
<td>Back problems</td>
<td>Back pain\textsuperscript{23}</td>
</tr>
<tr>
<td>Hyperlipidemia</td>
<td>Hyperlipidemia\textsuperscript{24}</td>
</tr>
</tbody>
</table>

\textsuperscript{a} In terms of public and private expenditure for 2008.\textsuperscript{13}
\textsuperscript{b} As selected by authors of the present study.

Results

The Agency for Healthcare Research and Quality\textsuperscript{13} listed the following 10 conditions as the costliest: heart disease, cancer, mental disorders, trauma-related disorders, osteoarthritis, chronic obstructive pulmonary disease/asthma, hypertension, diabetes, back problems, and hyperlipidemia. The corresponding Wikipedia articles\textsuperscript{15-24} are listed in Table 1. Examples of the descriptive terms we used to categorize the findings of each reviewer are listed on Table 2.

Reviewers found a statistically significant discordance between Wikipedia and peer-reviewed sources for assertions that were similar ($P<.05$) in all but 1 of the conditions: trauma-related disorders (ie, concussions). The same was true for all assertions found by the blinded reviewers of the articles ($P<.05$ for all conditions except concussions). In 4 articles—major depressive disorder, osteoarthritis, chronic obstructive pulmonary...
an assertion, whereas another might have assumed the statement to be common knowledge and erroneously not recorded it as an assertion. These incongruent criteria for assertions may explain the difference found between reviewers.

Although 9 of 10 articles demonstrated discordance between Wikipedia articles and the peer-reviewed sources, the article on concussions did not. This finding may have occurred because Wikipedia has a number of different contributors to each article and the contributors to this particular article were more expert.

The present study had 5 main limitations. First, it did not address errors of omission, but rather was designed to detect assertional errors. It is possible that the Wikipedia article did not contain important information about a topic. However, we opted not to examine errors of omission because of the subjectivity involved with determining what should be included in a review article on a specific medical topic. Second, the present study would have been stronger if more than 2 reviewers were assigned to each article. A future study design could use additional reviewers with more varied specializations to strengthen its findings. Third, we used any peer-reviewed reference as a standard that included an initial search through a subscription-only service (UpToDate). Fourth, we used physicians-in-training rather than content experts as reviewers, which may have created a bias that the present study was not designed to measure. Lastly, we did not check the assertions in the peer-reviewed sources, a limitation that may prove important because peer-reviewed sources are often not in agreement. Future studies might also include how the convenience of Wikipedia may influence perception of the reliability of the information found.

Conclusion

Most Wikipedia articles for the 10 costliest conditions in the United States contain errors compared with standard peer-reviewed sources. Health care professionals, trainees, and patients should use caution when using Wikipedia to answer questions regarding patient care.

Our findings reinforce the idea that physicians and medical students who currently use Wikipedia as a medical reference should be discouraged from doing so because of the potential for errors.
### Table 3.
No. of Similar and Dissimilar Assertions and Corresponding *P* Values of 10 Wikipedia Articles

<table>
<thead>
<tr>
<th>Wikipedia Article</th>
<th>Similar</th>
<th>Dissimilar</th>
<th>Both</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Concordance</td>
<td>Discordance</td>
<td>Concordance</td>
<td>Discordance</td>
</tr>
<tr>
<td>Lung Cancer</td>
<td>73</td>
<td>27</td>
<td>31</td>
<td>17</td>
</tr>
<tr>
<td>Reviewer 1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Reviewer 2</td>
<td>83</td>
<td>18</td>
<td>17</td>
<td>2</td>
</tr>
<tr>
<td><em>P</em> value</td>
<td>&lt;.001</td>
<td>.99</td>
<td>.001</td>
<td>.001</td>
</tr>
<tr>
<td>Diabetes Mellitus</td>
<td>37</td>
<td>1</td>
<td>15</td>
<td>3</td>
</tr>
<tr>
<td>Reviewer 1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Reviewer 2</td>
<td>34</td>
<td>2</td>
<td>40</td>
<td>7</td>
</tr>
<tr>
<td><em>P</em> value</td>
<td>&lt;.001</td>
<td>&lt;.001</td>
<td>&lt;.001</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Osteoarthritis</td>
<td>33</td>
<td>8</td>
<td>9</td>
<td>4</td>
</tr>
<tr>
<td>Reviewer 1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Reviewer 2</td>
<td>33</td>
<td>8</td>
<td>19</td>
<td>13</td>
</tr>
<tr>
<td><em>P</em> value</td>
<td>.001</td>
<td>.003</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Coronary Artery Disease</td>
<td>17</td>
<td>7</td>
<td>24</td>
<td>4</td>
</tr>
<tr>
<td>Reviewer 1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Reviewer 2</td>
<td>19</td>
<td>9</td>
<td>8</td>
<td>5</td>
</tr>
<tr>
<td><em>P</em> value</td>
<td>.029</td>
<td>.388</td>
<td>.012</td>
<td></td>
</tr>
<tr>
<td>Chronic Obstructive Pulmonary Disease</td>
<td>36</td>
<td>16</td>
<td>8</td>
<td>3</td>
</tr>
<tr>
<td>Reviewer 1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Reviewer 2</td>
<td>63</td>
<td>10</td>
<td>24</td>
<td>3</td>
</tr>
<tr>
<td><em>P</em> value</td>
<td>&lt;.001</td>
<td>&lt;.001</td>
<td>&lt;.001</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Hyperlipidemia</td>
<td>17</td>
<td>0</td>
<td>11</td>
<td>0</td>
</tr>
<tr>
<td>Reviewer 1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Reviewer 2</td>
<td>19</td>
<td>4</td>
<td>4</td>
<td>2</td>
</tr>
<tr>
<td><em>P</em> value</td>
<td>&lt;.001</td>
<td>.375</td>
<td>.001</td>
<td></td>
</tr>
<tr>
<td>Concussion</td>
<td>40</td>
<td>24</td>
<td>22</td>
<td>26</td>
</tr>
<tr>
<td>Reviewer 1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Reviewer 2</td>
<td>26</td>
<td>8</td>
<td>21</td>
<td>3</td>
</tr>
<tr>
<td><em>P</em> value</td>
<td>.888</td>
<td>.56</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hypertension</td>
<td>27</td>
<td>13</td>
<td>29</td>
<td>11</td>
</tr>
<tr>
<td>Reviewer 2</td>
<td>62</td>
<td>12</td>
<td>7</td>
<td>0</td>
</tr>
<tr>
<td><em>P</em> value</td>
<td>&lt;.001</td>
<td>.481</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Major Depressive Disorder</td>
<td>36</td>
<td>9</td>
<td>20</td>
<td>7</td>
</tr>
<tr>
<td>Reviewer 1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Reviewer 2</td>
<td>48</td>
<td>31</td>
<td>45</td>
<td>48</td>
</tr>
<tr>
<td><em>P</em> value</td>
<td>&lt;.001</td>
<td>&lt;.001</td>
<td>&lt;.001</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Back Pain</td>
<td>34</td>
<td>2</td>
<td>36</td>
<td>8</td>
</tr>
<tr>
<td>Reviewer 1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Reviewer 2</td>
<td>29</td>
<td>2</td>
<td>13</td>
<td>2</td>
</tr>
<tr>
<td><em>P</em> value</td>
<td>&lt;.001</td>
<td>.383</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

* Concordance or discordance found between each blinded reviewer for assertions that he or she found to be similar, dissimilar, or both. *P* values were calculated using the McNemar test for concordance and represent the ratings of 2 researchers who did not participate in the original review process and who tallied the assertions that were found by all blinded reviewers. The terms *assertions*, *similar*, *dissimilar*, *concordance*, and *discordance* are defined in Table 2.
References


