Is Wikipedia a reliable learning resource for medical students? Evaluating respiratory topics

Samy A. Azer
Curriculum Development and Research Unit, College of Medicine, King Saud University, Riyadh, Saudi Arabia

Submitted 13 August 2014; accepted in final form 7 November 2014

Azer SA. Is Wikipedia a reliable learning resource for medical students? Evaluating respiratory topics. Adv Physiol Educ 39: 5–14, 2015; doi:10.1152/advan.00110.2014.—The aim of the present study was to critically evaluate the accuracy and readability of English Wikipedia articles on the respiratory system and its disorders and whether they can be a suitable resource for medical students. On April 27, 2014, English Wikipedia was searched for articles on respiratory topics. Using a modified DISCERN instrument, articles were independently scored by three assessors. The scoring targeted content accuracy, frequency of updating, and quality of references. The readability of articles was measured using two other instruments. The mean DISCERN score for the 40 articles identified was 26.4 ± 6.3. Most articles covered causes, signs and symptoms, prevention, and treatment. However, several knowledge deficiencies in the pathogenesis of diseases, investigations needed, and treatment were observed. The total number of references for the 40 articles was 1,654, and the references varied from 0 to 168 references, but several problems were identified in the list of references and citations made. The readability of articles was in the range of 9.4 ± 1.8 to 22.6 ± 10.7 using the Flesch-Kincaid Grade Level instrument and 10.0 ± 2.6 to 19.6 ± 8.3 using the Readability Coleman-Liau index. A strong correlation was found between the two instruments \( r^2 = 0.744, P < 0.001 \). The agreement between the assessors had mean \( \kappa \) scores in the range of 0.712–0.857. In conclusion, despite the effort placed in creating engines, including Google and Yahoo, usually show Wikipedia articles written and edited anonymously by volunteers. This compromise information credibility for speed of use, accessibility, and convenience when selecting their learning sources (27, 43).

The main criticism directed against Wikipedia articles, at least for scientific and medical articles, is related to the paucity of formal editorial services of such resources and the fact that articles are written and edited anonymously by volunteers. This has created significant skepticism around the reliability and validity of its content (18). While a few empirical studies have evaluated the medical content of Wikipedia (19), there are no studies in the literature assessing the scientific accuracy and readability of articles that could be used by medical students when they search Wikipedia for their learning issues. Also, there are no studies assessing accuracy of content and readability of Wikipedia articles covering body systems such as the respiratory system. Such studies may be of value to medical educators and course designers as well as to medical students.
Considering the tendency of medical students to prefer online resources such as Wikipedia in their search for information, the present study was designed with the aim to critically evaluate the content, accuracy of information, quality of references, frequency of updating, and readability of English Wikipedia articles on the respiratory system and its common disorders.

METHODS

Search strategy. English Wikipedia articles on the respiratory system and its disorders were searched on April 27, 2014. The selection of topics was based on careful analysis of the chapters on respiratory diseases of five medical textbooks commonly prescribed in undergraduate medical programs (Table 1) and eMedicine (Medscape) topics on respiratory diseases (http://www.medscape.com). Being written by experts in the field and regularly reviewed and updated, these resources were used as a reference in the selection of topics and in the assessment of the Wikipedia articles. These textbooks were selected because they are recommended in most medical schools worldwide and have been reviewed in general and internal medicine journals such as the British Medical Journal, Journal of the American Medical Association, and New England Journal of Medicine (4, 14, 29, 46). Articles on eMedicine (Medscape) were authored by experts and are regularly updated on a yearly basis.

Another source for identifying the respiratory topics was the M1 and M2 curricula. This was based on the author’s long experience in medical school curricula at the University of Sydney and University of Melbourne in Australia as well as other detailed curricula for undergraduate medical schools such as curriculum verification reports (University of California-San Francisco School of Medicine; http://www.aamc.org/download/363632/data/sampleverifiereptucsf.pdf). Although M1 and M2 curriculum content often vary among medical schools, the focus in the search was about common core content.

The search strategy consisted of the following: 1) targeting articles covering common disorders affecting the respiratory system, 2) covering specific topics as well as general topics on the respiratory system that may be needed during the introduction of the module/block on the respiratory system, and 3) articles about rare conditions/syndromes related to the respiratory system were not included. The process of identifying topics from the five textbooks and eMedicine website was based on the degree of emphasis given by each resource, being considered in M1 and M2 curricula, and being covered in core contents required in undergraduate medical education. This process enabled identification of the topics. Topics that were given priority in these resources were identified by three assessors: the author (who is a medical consultant, a professor of medical education, and worked in curriculum development for over 15 years) and two research assistants (both with medical backgrounds). The process of identifying topics and searching for Wikipedia articles and their evaluation is shown in Fig. 1.

A total of 40 articles in the English language were finally identified, and copies of each article were obtained on the same day. The aim of making copies to assessors rather than accessing the actual electronic version is to ensure that the three assessors used the same version. This is particularly important as Wikipedia articles are frequently updated and changed.

Evaluation tool. A modified DISCERN instrument was used in the evaluation of accuracy of Wikipedia articles. The original DISCERN instrument was developed at Oxford University and consisted of 16 questions, including an overall evaluation question (22). It was originally developed to evaluate the quality of written consumer health information (48). This tool has been used in evaluating the quality of internet sites providing information relating to oral cancer (28), information resources on treatment for asthma and atopic dermatitis (5), quality of Web-based health information resources (30), quality and safety of Web-based information about herbal medicines in the treatment of cancer (34), colorectal cancer information on the internet (15), and the quality of snoring treatment information on the internet (49).

However, the original DISCERN instrument does not address the quality of illustrations, images, and tables included in an online resource. It also does not assess if there are gaps or deficiencies in the information provided. Therefore, it was decided to modify the instrument to enable the evaluation of Wikipedia articles. The modified DISCERN instrument is composed of 10 questions covering the depth of content, scientific accuracy, completeness, justification/evidence given, and readability grade (see the Supplemental Material for the APPENDIX).1 The instrument was tested before being used in this study (3).

The scoring system follows the scoring used in the DISCERN instrument: 1 was given if the answer was “no,” 2 was given if the answer was “partially” yes, and 5 was given if the answer was “yes.” For the last question 1 was given if there were “serious or extensive shortcomings,” 3 was given if there were “potentially important but not serious shortcomings,” and 5 was given if there were “minimal shortcomings.”

Rank 1, “serious or extensive shortcomings,” is defined as finding serious scientific errors and significant deficiencies in key principles and details related to a topic. Rank 3 is defined as a lack of clarity in information provided, finding confusing or contradictory statements, absence of evidence, and/or absence of one of the key principles related to a topic. Rank 5, “minimal shortcomings,” is defined as finding the topic well presented and scientifically correct but lacking minor issues that will not affect the educational usefulness and validity of information given. An error was recorded when two or more evaluators agreed that an article presented a particular piece of information that was disputatious or in disagreement with current medical textbooks (3).

Piloting the use of the instrument. The aims of piloting the use of the instrument before evaluating Wikipedia articles were to 1) orient that the three assessors to the instrument and how to use it, 2) ensure that the assessors were able to apply each item correctly, and 3) maximize the degree of agreement among the three assessors. To conduct the piloting portion, a total of 10 Wikipedia articles (other than those included in the study) were used. The three assessors were asked to apply the modified DISCERN instrument independently. The scores were then discussed in a meeting. None of the assessors shared their scores with other assessors before the meeting. Differences in the scores allocated were discussed and the reasons for inconsistencies were discussed until a consensus was reached. This exercise helped the evaluators on how to apply the instrument and assess the different

Table 1. The five medical textbooks used as references in evaluating Wikipedia articles


1 Supplemental Material for this article is available at the Advances in Physiology Education website.
components of the tool. Another 10 articles were evaluated independently in the same way discussed. The results showed an agreement among the three evaluators of 80–90%, indicating optimum agreement and readiness to use the instrument.

Assessing the respiratory articles. The 40 respiratory Wikipedia articles were then assessed independently by the 3 assessors using the modified DISCERN instrument. The findings were placed on a Microsoft Excel 2010 document (Microsoft, Redmond, WA), and the outcomes were discussed in a meeting.

Assessing the references. For each article, the total number of references was recorded. References were also categorized into the following subcategories: peer-reviewed journals, textbooks, educational guidelines/position statements, reports/editorials, professional websites, general websites, and news/media/others. References for peer-reviewed journals were further subdivided according to the year of publication into those published in the last 5 yr and those published earlier.

Citations were checked for accuracy and whether what was stated in the article matched with the citation provided. In addition, references were checked in regard to 1) consistency in the way they were presented; 2) use of capital and small letters, italics, and bold; 3) use of appropriate international abbreviations or full journal titles; and 4) any missing information or errors in the references.

Frequency of updating articles. The “view history” button on the top of each article and the subsequent links, such as “review history statistics” and “edit history,” were carefully examined for the frequency of information updating, including total number of revisions, average time between edits, average edits per month, and revisions in the last 12 mo for each article.

Assessment of readability. The aim for assessing readability was to examine the reading level and whether the articles were suited for college students. Two methods were used to calculate readability level: the validated Flesch-Kincaid Grade Level and Coleman-Liau index (17). The Flesch-Kincaid Grade Level test is designed to indicate comprehension difficulty when reading a passage of an academic article in English. The methods depend on measurements of word and sentence length but have different weighting factors. The Flesch-Kincaid Grade Level formula is as follows:

\[
0.39 \left( \frac{\text{Total words}}{\text{Total sentences}} \right) + 11.8 \left( \frac{\text{Total syllables}}{\text{Total words}} \right) - 15.59
\]

The result is a number corresponding with a United States grade level. The Coleman-Liau index relies on characters rather than syllables per word. Although opinion varies in regard to its accuracy, characters are more readily counted by the computer program. The Coleman-Liau index can be calculated using the following formula:

\[
\text{Coleman – Liau Index} = 0.0588L - 0.296S - 15.8
\]

where \( L \) is the average number of letters per 100 words and \( S \) is the average number of sentences per 100 words. To measure the readability of each article, 3 passages of text ranging between 150–600 words in length were randomly selected from the beginning, middle, and end. The Flesch-Kincaid Grade Level and Coleman-Liau index were calculated using a free program provided by Readability Formulas (http://www.readabilityformulas.com/freetests/six-readability.formula.php). To conduct the calculation, headings and references that could artificially affect the calculation were omitted. Mean ± SD readability scores for each article were calculated.

Statistical analysis. Scores obtained from each assessor were entered on a Microsoft Excel 2010 document and were checked by two assessors before any analysis was conducted. Analysis was conducted with SPSS software (version 18.0 for MS Windows, SPSS, Chicago, IL). Findings are reported as means ± SD for DISCERN scores obtained for each Wikipedia topic and readability scores. Correlation between the two reading scores using the two methods and between

Fig. 1. Search process and evaluation of Wikipedia articles on respiratory topics.
## Table 2. Accuracy scores, numbers of images, illustrations, and tables, and readability scores of each Wikipedia respiratory article included in the study

<table>
<thead>
<tr>
<th>Topic Title</th>
<th>Number of Pages</th>
<th>Accuracy Score (means ± SD)</th>
<th>Number of Images/Illustrations</th>
<th>Number of Audio Files</th>
<th>Number of Tables</th>
<th>Total</th>
<th>Readability Fleisch-Kincaid Grade Level (means ± SD)</th>
<th>Readability Coleman-Liau Index (means ± SD)</th>
<th>External Links/ Further Reading</th>
</tr>
</thead>
<tbody>
<tr>
<td>Acute bronchitis</td>
<td>4</td>
<td>20.67 ± 0.57</td>
<td>1 Illustration</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>12.80 ± 1.83</td>
<td>12.33 ± 0.50</td>
<td>2 External links</td>
</tr>
<tr>
<td>Acute respiratory distress syndrome</td>
<td>13</td>
<td>25.00 ± 1.00</td>
<td>1 X-ray, 1 histopathological image, and 1 illustration</td>
<td>0</td>
<td>0</td>
<td>3</td>
<td>15.70 ± 3.25</td>
<td>16.67 ± 3.05</td>
<td>3 External links</td>
</tr>
<tr>
<td>Alveolar lung disease</td>
<td>1</td>
<td>15.67 ± 0.57</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>16.43 ± 2.55</td>
<td>16.33 ± 2.51</td>
<td>0</td>
</tr>
<tr>
<td>Asbestosis</td>
<td>7</td>
<td>26.33 ± 1.15</td>
<td>7 Histopathological images and 2 X-rays</td>
<td>0</td>
<td>0</td>
<td>9</td>
<td>15.30 ± 3.12</td>
<td>15.00 ± 2.00</td>
<td>7 External links</td>
</tr>
<tr>
<td>Asthma</td>
<td>25</td>
<td>35.33 ± 2.08</td>
<td>4 Photos, 1 histopathological image, and 2 maps (epidemiology)</td>
<td>1 Sound of wheezing</td>
<td>0</td>
<td>11</td>
<td>12.33 ± 1.00</td>
<td>12.00 ± 1.00</td>
<td>1 External link and 3 further readings</td>
</tr>
<tr>
<td>Breathing (mechanisms of breathing)</td>
<td>6</td>
<td>20.33 ± 0.57</td>
<td>1 X-ray and 1 illustration</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>12.16 ± 1.05</td>
<td>12.67 ± 1.15</td>
<td>1 External link</td>
</tr>
<tr>
<td>Bronchiectasis</td>
<td>7</td>
<td>23.00 ± 1.73</td>
<td>1 Histopathological image, 1 pathological image, and 2 illustrations</td>
<td>0</td>
<td>0</td>
<td>4</td>
<td>15.76 ± 0.75</td>
<td>17.33 ± 0.57</td>
<td>3 External links</td>
</tr>
<tr>
<td>Chest pain</td>
<td>5</td>
<td>18.67 ± 0.57</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>17.7 ± 2.53</td>
<td>15.00 ± 3.60</td>
<td>1 External link</td>
</tr>
<tr>
<td>Chronic obstructive pulmonary disease</td>
<td>27</td>
<td>37.33 ± 2.88</td>
<td>2 Pathological images, 2 maps (epidemiology), 2 illustrations, 4 X-rays, and 1 CT image</td>
<td>1 Sound of wheezing</td>
<td>0</td>
<td>14</td>
<td>9.46 ± 1.80</td>
<td>10.00 ± 2.64</td>
<td>1 External link and 3 further readings</td>
</tr>
<tr>
<td>Clubbing</td>
<td>1</td>
<td>14.67 ± 0.57</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>16.76 ± 2.33</td>
<td>13.33 ± 0.57</td>
<td>0</td>
</tr>
<tr>
<td>Cough</td>
<td>7</td>
<td>25.00 ± 1.73</td>
<td>1 Photo and 1 illustration</td>
<td>0</td>
<td>0</td>
<td>3</td>
<td>11.80 ± 2.00</td>
<td>13.33 ± 2.08</td>
<td>1 External link</td>
</tr>
<tr>
<td>Cystic fibrosis</td>
<td>26</td>
<td>36.00 ± 1.00</td>
<td>3 Photos and 5 illustrations</td>
<td>0</td>
<td>1</td>
<td>9</td>
<td>14.20 ± 0.91</td>
<td>13.33 ± 0.57</td>
<td>6 External links and 11 further readings</td>
</tr>
<tr>
<td>Diffuse panbronchiolitis</td>
<td>10</td>
<td>34.67 ± 2.51</td>
<td>2 CT images and 2 illustrations</td>
<td>0</td>
<td>0</td>
<td>4</td>
<td>14.26 ± 0.73</td>
<td>12.67 ± 2.08</td>
<td>0</td>
</tr>
<tr>
<td>Dyspnea</td>
<td>8</td>
<td>27.33 ± 2.08</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>14.60 ± 1.56</td>
<td>15.67 ± 0.57</td>
<td>2 External links</td>
</tr>
<tr>
<td>Gas exchange</td>
<td>4</td>
<td>20.67 ± 0.57</td>
<td>1 Illustration</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>13.40 ± 1.25</td>
<td>13.00 ± 1.00</td>
<td>2 External links</td>
</tr>
<tr>
<td>Granulomatosis with polyangiitis</td>
<td>6</td>
<td>26.67 ± 2.64</td>
<td>1 Histopathological image and 2 photos</td>
<td>0</td>
<td>0</td>
<td>3</td>
<td>18.43 ± 7.17</td>
<td>16.67 ± 2.88</td>
<td>2 External links</td>
</tr>
<tr>
<td>Hemoptysis</td>
<td>3</td>
<td>23.33 ± 0.57</td>
<td>1 Illustration</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>15.30 ± 1.68</td>
<td>17.33 ± 1.15</td>
<td>7 External links</td>
</tr>
<tr>
<td>Idiopathic pulmonary fibrosis</td>
<td>6</td>
<td>28.00 ± 1.73</td>
<td>2 Illustrations, 1 photo, 1 X-ray, 1 CT image, and 2 histopathological images</td>
<td>1 Sound of crackles</td>
<td>0</td>
<td>8</td>
<td>12.46 ± 2.87</td>
<td>10.67 ± 4.16</td>
<td>1 External link</td>
</tr>
<tr>
<td>Interstitial lung disease</td>
<td>6</td>
<td>25.67 ± 0.57</td>
<td>1 Illustration, 1 histopathological image, 1 pathological image, and 2 X-rays</td>
<td>0</td>
<td>0</td>
<td>5</td>
<td>22.63 ± 10.79</td>
<td>19.67 ± 8.32</td>
<td>13 External links</td>
</tr>
<tr>
<td>Lung abscess</td>
<td>4</td>
<td>23.67 ± 0.57</td>
<td>1 CT image and 1 pathological image</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>17.43 ± 5.05</td>
<td>18.00 ± 3.60</td>
<td>0</td>
</tr>
<tr>
<td>Lung cancer</td>
<td>24</td>
<td>35.67 ± 1.52</td>
<td>1 X-ray, 1 CT image, 2 illustrations, 4 pathological images, and 2 maps (epidemiology)</td>
<td>0</td>
<td>3</td>
<td>13</td>
<td>12.10 ± 3.03</td>
<td>11.00 ± 1.73</td>
<td>1 External link</td>
</tr>
<tr>
<td>Mesothelioma</td>
<td>19</td>
<td>32.00 ± 1.73</td>
<td>1 X-ray, 2 CT images, 1 microbiological image, and 2 pathological images</td>
<td>0</td>
<td>1</td>
<td>7</td>
<td>17.03 ± 4.00</td>
<td>11.67 ± 0.57</td>
<td>14 External links</td>
</tr>
<tr>
<td>Obstructive sleep apnea</td>
<td>16</td>
<td>30.33 ± 0.57</td>
<td>1 Illustrations</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>14.73 ± 0.75</td>
<td>12.33 ± 0.57</td>
<td>3 External links and 5 further readings</td>
</tr>
<tr>
<td>Occupational lung disease</td>
<td>1</td>
<td>17.67 ± 0.57</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>16.66 ± 1.80</td>
<td>16.33 ± 2.08</td>
<td>0</td>
</tr>
<tr>
<td>Pleural effusion</td>
<td>9</td>
<td>26.33 ± 0.57</td>
<td>3 X-rays, 1 CT image, 1 ultrasound, 1 microbiological image, and 1 illustration</td>
<td>0</td>
<td>1</td>
<td>8</td>
<td>15.06 ± 5.31</td>
<td>12.67 ± 2.08</td>
<td>3 External links</td>
</tr>
<tr>
<td>Pleurisy</td>
<td>12</td>
<td>29.67 ± 0.57</td>
<td>3 Illustrations</td>
<td>0</td>
<td>0</td>
<td>3</td>
<td>15.5 ± 10.33</td>
<td>12.33 ± 4.04</td>
<td>2 External links</td>
</tr>
<tr>
<td>Pneumonia</td>
<td>21</td>
<td>30.33 ± 1.15</td>
<td>2 X-rays, 1 CT scan, 1 microbiological image, 3 illustrations, 1 map (epidemiology)</td>
<td>1 Sound of crackles</td>
<td>2</td>
<td>11</td>
<td>14.70 ± 2.36</td>
<td>13.33 ± 0.57</td>
<td>1 External link</td>
</tr>
<tr>
<td>Pneumothorax</td>
<td>15</td>
<td>33.00 ± 1.73</td>
<td>2 X-rays, 1 CT image, and 2 illustrations</td>
<td>0</td>
<td>1</td>
<td>6</td>
<td>14.36 ± 0.61</td>
<td>14.33 ± 2.30</td>
<td>0</td>
</tr>
<tr>
<td>Pulmonary embolism</td>
<td>15</td>
<td>31.33 ± 1.15</td>
<td>1 X-ray, 2 CT images, 1 angiogram, 1 scan, 1 ECG, and 3 illustrations</td>
<td>0</td>
<td>0</td>
<td>9</td>
<td>15.06 ± 5.42</td>
<td>13.33 ± 4.16</td>
<td>0</td>
</tr>
<tr>
<td>Pulmonary function testing</td>
<td>6</td>
<td>23.67 ± 1.15</td>
<td>1 Illustration</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>17.13 ± 1.24</td>
<td>16.33 ± 2.08</td>
<td>0</td>
</tr>
<tr>
<td>Pulmonary surfactant</td>
<td>7</td>
<td>24.33 ± 1.15</td>
<td>2 Illustrations</td>
<td>0</td>
<td>1</td>
<td>3</td>
<td>12.56 ± 0.85</td>
<td>13.67 ± 1.15</td>
<td>1 External link</td>
</tr>
</tbody>
</table>

Continued
the DISCERN scores and other variables (e.g., number of revisions, total number of pages, and number of authors/editors) were also studied (11, 44). The interrater analysis between evaluators for each item in the modified DISCERN instrument was assessed by calculating Cohen’s $\kappa$ range of score (35, 40).

**RESULTS**

**Depth and accuracy scores.** Table 2 shows accuracy scores (means ± SD) calculated using the modified DISCERN instrument, number of images, illustrations, and tables. The minimum score was 14.6 for the article on clubbing, and the maximum score was 38.3 for the article on tuberculosis. Of the 40 articles, 13 articles scored 30 or higher, 20 articles scored 20 to 29, and 7 articles scored 14–19. The number of pages ranged from 1 to 27 pages, indicating that articles varied significantly in regard to the size of the topic covered and details given. For example, the article on chronic obstructive pulmonary disease was covered in 27 pages, had 127 references, and had 4 external links and further readings. The article on cystic fibrosis was covered in 26 pages, had 136 references, and had 6 external links and 11 further readings. Incomplete articles were covered on one page, such as the articles on alveolar lung disease, clubbing, and occupational lung disease. These articles had no references or external links. The number of images/illustrations varied from 0 to 16, and the number of tables varied from 0 to 3.

Although articles followed the template created by Wikipedia for medical articles, some articles were incomplete and most articles were deficient in discussing the pathogenesis and mechanisms of diseases. In addition, there were deficiencies in clinical pictures, investigations, and management of respiratory diseases. The time required to complete the evaluation of each topic varied from 15 to 150 min.

**Article references.** Table 3 shows an analysis of the references included in the Wikipedia respiratory articles. The total number of references in the 40 articles was 1,654, and the number of references varied from 0 to 168 references (41.3 ± 48.2, mean ± SD). The majority of references were from peer-reviewed medical and scientific journals, making a total of 1,045 (63%). Other references were guidelines/position statements ($n = 68, 4\%$), textbooks ($n = 268, 16\%$), reports/editorials ($n = 31, 2\%$), professional websites ($n = 123, 7\%$), general websites ($n = 65, 4\%$), and news/media and others ($n = 54, 3\%$).

The number of citations from peer-reviewed journals published in the last 5 yr was only 312 (19\%). Common problems in citations and the list of references were observed in most articles. These errors/problems can be summarized as follows: 1) citation of wrong references and failing to cite appropriate references, 2) incomplete references, 3) inconsistencies in the way that references were presented (for example, the full title of the journal was stated and in other places the abbreviated title was given, the use of capital and small letters, the use of italics and bold lines and position statements produced by the respiratory societies and associations, such as the American Association for Respiratory Care (https://www.aarc.org), American Thoracic Society (http://www.thoracic.org), European Respiratory Society (http://www.ersnet.org), British Thoracic Society (https://www.brit-thoracic.org.uk), and Thoracic Society of Australia and New Zealand (http://www.thoracic.org.au).

**Article history and frequency of revisions.** Table 4 shows the article history, date created, number of revisions, number of authors, and frequency of revisions. While the earliest date for article creation was November 13, 2001, for the cystic fibrosis article, the most recent article was pulmonary function testing, created on March 19, 2008.

The total number of revisions for the 40 articles was 52,837 (1,320.9 ± 1,812.2, mean ± SD), and the total number of authors was 24,723 (618.1 ± 783.9, mean ± SD). The average time between edits varied from 1 to 104 days, and the average edits per month varied from 0 to 45 (9.8 ± 12.3, mean ± SD).
The revisions in the last 12 mo were 3,032 in total (75.8 ± 137.9, mean ± SD).

There was a significant correlation between the number of revisions and DISCERN scores ($r^2 = 0.659$, $P < 0.001$) and between the total number of revisions and number of authors/editors ($r^2 = 0.615$, $P < 0.001$) and between the number of revisions and total number of pages ($r^2 = 0.770$, $P < 0.001$).

**Article readability.** Two methods were used in calculating readability. Table 2 shows the readability scores (mean ± SD) of each article calculated by Flesch-Kincaid Grade Level and Readability Coleman-Liau index. For the first method, the readability was in the range of 9.4 ± 1.8–22.6 ± 10.7, whereas the second method recorded a range of 10.0 ± 2.6–19.6 ± 8.3. A strong correlation was found between the scores calculated by the two methods ($r^2 = 0.744$, $P < 0.001$). The scores obtained from the two methods indicate that Wikipedia articles were geared to a college level rather than to the public.

**Agreement between assessors.** Table 5 shows the agreement between assessors for each item included in the modified DISCERN instrument. Overall $\kappa$ scores were in the range of 0.712–0.857.

**DISCUSSION**

The aim of the present study was to evaluate the content, accuracy of information, quality of references, frequency of updating, and readability of English Wikipedia articles on the respiratory system and its common disorders. The findings show that of 40 articles, 13 articles scored 30 or higher, 20 articles scored 20 or 29, and 7 articles scored 14–19. The minimum score was 14.6, and the maximum score was 38.3. Although articles appeared to be well cited and reviewed regularly, several deficiencies in knowledge were identified. Several articles were incomplete, and articles were, in general, brief in linking knowledge with clinical applications, discussing the pathogenesis of diseases at body system, organ, cellu-
Several studies have evaluated Wikipedia articles as a source of information for pharmacists (7), medical students (9, 21, 37), nursing and health professional students (16, 45), college students (27), junior physicians (20), and patients (22). A study by Laurent and Vickers (26) showed that Wikipedia was ranked among the first 10 results in 71–85% of search engines (Google, Google UK, Yahoo, and MSN). In self-directed learning resources such as the scientific content of each article and accuracy of information given compared with the information provided by five standard medical textbooks and eMedicine (Medscape) resources. Other parameters

<table>
<thead>
<tr>
<th>Number</th>
<th>Topic Title</th>
<th>Date Created</th>
<th>Latest Edit</th>
<th>Total Number of Revisions</th>
<th>Total Number of Distinct Authors</th>
<th>Average Time Between Edits, days</th>
<th>Average Edits Per Month</th>
<th>Revision in Last 12 mo</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Acute bronchitis</td>
<td>23 August 2003</td>
<td>24 April 2014</td>
<td>1061</td>
<td>701</td>
<td>3.67</td>
<td>8.28</td>
<td>18</td>
</tr>
<tr>
<td>2</td>
<td>Acute respiratory distress syndrome</td>
<td>22 February 2004</td>
<td>20 April 2014</td>
<td>528</td>
<td>287</td>
<td>7.02</td>
<td>4.33</td>
<td>89</td>
</tr>
<tr>
<td>3</td>
<td>Alveolar lung disease</td>
<td>13 February 2006</td>
<td>22 March 2014</td>
<td>33</td>
<td>25</td>
<td>89.67</td>
<td>0.34</td>
<td>3</td>
</tr>
</tbody>
</table>
| 4      | Asth...
Table 5. Summary of agreement between evaluators using Cohen κ score

<table>
<thead>
<tr>
<th>Question</th>
<th>Evaluator 1</th>
<th>Evaluator 2</th>
<th>Evaluator 3</th>
<th>Reviewer Variability (κ Range)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Are the aims of the topic clearly stated at the beginning of the article?</td>
<td>2.28 (2.07–2.48)</td>
<td>1.88 (1.69–2.06)</td>
<td>2.10 (1.90–2.30)</td>
<td>0.580–0.696</td>
</tr>
<tr>
<td>2. Does the article cover the needed subtitles and key concepts related to the topic?</td>
<td>2.55 (2.28–2.82)</td>
<td>2.35 (2.13–2.57)</td>
<td>2.50 (2.23–2.77)</td>
<td>0.684–0.924</td>
</tr>
<tr>
<td>3. Is the information provided throughout the article scientifically correct and in agreement with current valid resources and textbooks?</td>
<td>3.03 (2.80–3.25)</td>
<td>2.68 (2.48–2.87)</td>
<td>2.88 (2.66–3.09)</td>
<td>0.431–0.755</td>
</tr>
<tr>
<td>4. Is the article neutral and not based on personal views?</td>
<td>3.00 (2.82–3.18)</td>
<td>2.73 (2.56–2.89)</td>
<td>2.93 (2.74–3.11)</td>
<td>0.481–0.739</td>
</tr>
<tr>
<td>5. Is the article balanced?</td>
<td>2.78 (2.55–3.00)</td>
<td>2.53 (2.33–2.72)</td>
<td>2.68 (2.45–2.90)</td>
<td>0.741–0.750</td>
</tr>
<tr>
<td>6. Is it clear what sources of information were used to compile the publication (references and links to professional websites)?</td>
<td>2.85 (2.52–3.18)</td>
<td>2.70 (2.38–3.02)</td>
<td>2.75 (2.44–3.06)</td>
<td>0.862–0.931</td>
</tr>
<tr>
<td>7. Has the article been regularly updated and amended?</td>
<td>2.90 (2.55–3.25)</td>
<td>2.85 (2.51–3.19)</td>
<td>2.90 (2.55–3.25)</td>
<td>0.922–0.999</td>
</tr>
<tr>
<td>8. Are the key areas in the article completed and do not need further addition?</td>
<td>2.63 (2.28–2.97)</td>
<td>2.50 (2.17–2.83)</td>
<td>2.58 (2.23–2.92)</td>
<td>0.899–0.933</td>
</tr>
<tr>
<td>9. Do the images, figures, and tables provided in the article support the information provided and enhance understanding of points raised?</td>
<td>2.53 (2.21–2.84)</td>
<td>2.35 (2.08–2.62)</td>
<td>2.45 (2.14–2.76)</td>
<td>0.743–0.894</td>
</tr>
<tr>
<td>10. What is your overall rating of the whole article as a source of information to medical students?</td>
<td>2.98 (2.73–3.22)</td>
<td>2.80 (2.56–3.04)</td>
<td>2.83 (2.59–3.06)</td>
<td>0.733–0.961</td>
</tr>
</tbody>
</table>

considered in the evaluation were the number of images, illustrations, multimedia, and tables, the number of references, the number of peer-reviewed references, the frequency of article updating, and the number of authors/editors for each article. Finally, the measurement of readability of the articles used two methods, which showed a strong correlation between the scores calculated from each method. Both methods showed that the articles were geared to the college level.

Although most articles were regularly updated (52,837 times for the 40 articles), by a good number of authors/editors (24,723 for the 40 articles), and the majority of references were from peer-reviewed journals (63%), several problems were identified in the list of references of most articles and in the citations given. There were deficiencies in the use of educationally useful images/illustrations and tables to explain information given in depth. In addition, the use of multimedia was limited and not used in a way that enhanced the article. Most images and illustrations were not labeled, and the pathological changes were not identified/highlighted or compared with another image illustrating normal image. Although a good number of peer-reviewed journals were included in the references, in several situations these references did not target key papers related to the topic discussed and recent guidelines/position statements from societies/associations were not included. Although the number of editors/authors involved in the 40 articles looks impressive, there was no information available about these authors and their experience. Wikipedia allows anyone to edit and make changes to articles; thus, the authors are not necessarily doctors or respiratory physicians or academics specialized in an area related to the respiratory system (6, 18, 52). Also, the updating of the articles was usually about grammatical changes and editing sentences rather than making significant changes to the scientific content (47). For example, the papers titled “alveolar lung disease” and “clubbing” were only 1 page long each and were updated 33 and 25 times, respectively, indicating that most changes were limited to minor changes rather than expanding and adding new content. Although topics were created approximately at the same time (between 2002 and 2004), the numbers of revisions and edits were so different. This may be related to the lack of editing of Wikipedia articles to ensure harmony and a balanced distribution of contents in articles. It is also related to the interest of Wikipedians to update particular topics of interest to them (such as bronchial asthma and chronic obstructive pulmonary disease) rather than others.

However, the study is not free from limitations. The results from the study are based on the analysis of 40 English Wikipedia articles on the respiratory system and its disorders, and the results cannot be generalized to other articles on Wikipedia covering other body systems or other languages, which limits the generalization of the results. In addition, the findings from this study reflect the status of Wikipedia articles on respiratory diseases at the time of the study. Wikipedia articles inherently are in a constant state of influx, and it has been shown that articles improve over time (10). In this context, the present study has demonstrated a significant correlation between DISCERN scores and the number of updates over time.

As indicated by Wikipedia administrators, several articles were incomplete. These deficiencies can be categorized as follows: 1) articles in their early stages, for example, the articles titled “alveolar lung disease,” “chest pain,” “clubbing,” “gas exchange,” and “ventilation/perfusion scan”; 2) articles showing deficiencies in some content or need tables, images, illustrations, or media to make the message meaningful and enhance their educational value, for example, the articles titled “acute bronchitis,” “breathing (mechanisms of breathing),” and “chest pain”; and 3) articles requiring the addition of proper citations for some statements, for example, the articles titled “acute bronchitis,” “alveolar lung disease,” “clubbing,” and “pulmonary surfactant.”

Recently, WikiProject Medicine has been introduced, where people interested in medical and health content on Wikipedia can discuss topics, collaborate, or debate issues. The Wikipedia articles also have been categorized in regard to their status by administrators. Details about the Wikipedia categorization are
given in the following link: http://en.wikipedia.org/wiki/Category:FA-Class_medicine_articles.

To enhance the quality of articles, it was suggested that professional societies/associations could create task forces for reviewing Wikipedia articles and even give a statement of approval on articles that have been completed and reviewed at a satisfactory level (42). University students could take the task of completing Wikipedia articles under supervision (8). It was proposed that professional bodies could support members as individual contributors to Wikipedia articles in their areas of expertise (50). It was also suggested that clinicians could contribute by editing Wikipedia (24). These proposals aim at the involvement of a wider range of professionals in the writing and editing of articles so that any perceived conflicts of interest could be resolved on a scientific basis.

In conclusion, the findings from the present study show that English Wikipedia articles on the respiratory system and its disorders follow the standardized template created by Wikipedia for medical/health articles. Although the articles were regularly updated and well cited, most articles were deficient in knowledge, particularly in the pathogenesis of diseases, differential diagnosis, and discussions of management plans and treatment options. Most articles cited peer-reviewed journals, but there were several errors in the list of references and allocation of citations in the article. In addition, guidelines and position statements issued by international thoracic societies/associations were not cited. The readability scores calculated using two methods showed that the articles were geared at a college level rather than to the public. Although the articles were regularly updated and reviewed, the articles were not free from scientific errors, and more work is needed to add/improve quality images, illustrations, multimedia, and tables to enhance the educational content of these articles. Therefore, the Wikipedia articles on the respiratory system and its disorders are not suitable for medical students as learning resources. Medical students should become aware of these problems in the Wikipedia articles and medical educators should direct them to educational content of these articles. Therefore, the Wikipedia articles on the respiratory system and its disorders follow the standardized template created by Wikipedia for medical/health articles. Although the articles were regularly updated and well cited, most articles were deficient in knowledge, particularly in the pathogenesis of diseases, differential diagnosis, and discussions of management plans and treatment options. Most articles cited peer-reviewed journals, but there were several errors in the list of references and allocation of citations in the article. In addition, guidelines and position statements issued by international thoracic societies/associations were not cited. The readability scores calculated using two methods showed that the articles were geared at a college level rather than to the public. Although the articles were regularly updated and reviewed, the articles were not free from scientific errors, and more work is needed to add/improve quality images, illustrations, multimedia, and tables to enhance the educational content of these articles. Therefore, the Wikipedia articles on the respiratory system and its disorders are not suitable for medical students as learning resources. Medical students should become aware of these problems in the Wikipedia articles and medical educators should direct them to appropriate resources that have been written/editied by scholarly authors and are regularly updated and accurate.

ACKNOWLEDGMENTS

The author thank Dr. Sarah Azer (Monash University) for assistance. Also, the author thanks Diana Azer (RMIT University) and Dr. Lily Scott for assistance in this work and Philip Fely, Laniefel Mendoza (Peng), and Mary Hadicron for secretarial help.

GRANTS

This work was funded by the College of Medicine Research Center, Deanship of Scientific Research (King Saud University, Riyadh, Saudi Arabia).

DISCLOSURES

No conflicts of interest, financial or otherwise, are declared by the author(s).

AUTHOR CONTRIBUTIONS

Author contributions: S.A.A. conception and design of research; S.A.A. performed experiments; S.A.A. analyzed data; S.A.A. interpreted results of experiments; S.A.A. prepared figures; S.A.A. drafted manuscript; S.A.A. edited and revised manuscript; S.A.A. approved final version of manuscript.

REFERENCES


