An Examination of Health, Medical and Nutritional Information on the Internet: A Comparative Study of Wikipedia, WebMD and the Mayo Clinic Websites

Behdin Nowrouzi, Basem Gohar, Camille Smith,
Laurentian University, Sudbury, Ontario, Canada
Centre for Research in Occupational Health and Safety
Bx_nowrouzi@laurentian.ca, BX_Gohar@laurentian.ca,CRSmith@laurentian.ca

Behnam Nowrouzi-Kia, Rubaiya Khan, Alicia McDougall, Martyna Garbaczewska,
University of Toronto, Toronto, Ontario,
Faculty of Arts and Science
behnam.nowrouzi.kia@mail.utoronto.ca, rubaiya_alam@yahoo.com, aliciamcdougall@gmail.com, martyna.garbaczewska@.utoronto.ca

Shalini Sivathasan
Division of Mood and Anxiety Disorders, Centre for Addiction and Mental Health, Toronto, ON
shalini.sivathasan@nyu.edu

Keith Brewster,
University of British Columbia, Kelowna, British Columbia, Canada
School of Health and Exercise Sciences
aokbrewster@me.com

Lorraine Carter
Nipissing University, North Bay, Ontario, Canada
School of Nursing
lorrainec@nipissingu.ca

Abstract

To examine the scope, completeness, credibility, and readability of health, medical and nutritional information found on Wikipedia, WebMD, and Mayo Clinic websites.A total of ninety-two statements, across nine health categories, were formulated and used to assess the selected websites. Trained raters used a standardized search protocol, electronic logs and the nine-item tool to assess for scope, completeness, credibility, and readability of online material across the three websites. In terms of the scope, answers for 91.3% of the 92 general health statements were available on Wikipedia. WebMD (89.1%), and the Mayo Clinic (81.5%) followed respectively. The Flesch Reading Ease Score (FRES) was significantly higher for Wikipedia compared to WebMD and the Mayo Clinic websites (p<0.001). The Flesch-Kincaid Grade Level (FKGL) scores were also significantly higher for Wikipedia compared to those for WebMD and the Mayo Clinic websites (p<0.001). Sources supporting the general health statements were present for 96.0% of webpages on the Mayo Clinic site, 95.1% of webpages for Wikipedia, and 94.9% of webpages for WebMD. he study findings demonstrate the importance of aligning information and services for health with the skills and abilities of its recipients. As a result, these findings may be used to improve patient health literacy and consequently reduce health disparities. s a growing number of people use online sources to obtain health, medical, and nutritional information, it is important that this information be complete, comprehensive in scope, and available at a literacy level that is accessible to the general population..

Key Words: Wikipedia, WebMD, Mayo Clinic, Health & medical information, Health literacy

Background and significance

With the fast rise of the Internet in the past decade, research on the use of Internet sources for health, nutrition, and medical information is growing (Couper et al., 2010; Helft, 2008). Furthermore, the Internet has rapidly become a source of health information for consumers and health professionals (McMullan, 2006). The Pew Internet and American Life project reported that 72% of American Internet users state that they looked online for health information within the past year (Pew Internet: Health, 2013). Moreover, 52% of smartphone owners have used their devices to look up health or medical information. Of those respondents, 13% indicated that they had begun their

search for health-related information at websites that specialized in health information such as WebMD and Mayo Clinic (Pew Internet: Health, 2013). Two percent of respondents indicated that they started their research at a more general site, such as Wikipedia (Pew Internet: Health, 2013). Furthermore, 55% of Internet users surveyed looked online for information about a specific disease or medical problem; 43% looked information about a certain medical treatment or procedure; and 27% searched for information about weight loss or weight management in the past year (Pew Internet & American Life Project, 2014).

Since the inception of Wikipedia in 2001, it (http://www.wikipedia.org) has grown rapidly into one of the largest reference websites in the world, attracting 470 million unique visitors monthly and hosting over 4.5 million articles in English (Wikipedia, 2014). It is now the sixth most visited website on the Internet (Alexa - The Web Information Company, 2013). The articles cover a broad range of subjects, are accessible at no cost to the user, and content is provided by the online community (members are permitted to edit articles). WebMD (http://www.webmd.com) is the second most visited health site and is run by an American corporation that provides health information services. As an online American health giant, WebMD is reported to have a net worth of 2.5 billion dollars (R. Cohen, Elhadad, & Birk, 2013) The Mayo Clinic's health information websites (http://www.mayoclinic.com/health-information) are a platform where more than 3,300 physicians and researchers disseminate their expertise across a wide variety of health topics (Mayo Clinic, 2013).

Despite growing awareness that health consumers use the Internet to educate themselves in order to make their health and medical decisions, there have been few organized initiatives to evaluate the state of online health information (Volsky, Baldassari, Mushti, & Derkay, 2012b). Moreover, no reviews have identified studies that examine the quality of online information (Volsky et al., 2012b). Studies have reported that online health information is often above the expected reading ability of a significant proportion of the American population (Berland et al., 2001; Graber, D Alessandro, & Johnson-West, 2002). While health consumers should be empowered to evaluate online content for themselves, literacy levels are a major concern. For instance, The American Medical Association reported that one in three patients have basic or below basic health literacy (O'Reilly, 2012). The National Patient Safety Foundation reported that such literacy gaps have profound health and medical ramifications (National Patient Safety Foundation, 2011). Health consumers with poor literacy skills have much poorer health outcomes than those who can read well.

Given the contextual information offered here and a lack of research evaluating the scope, completeness, credibility, and readability of health, medical, nutritional information found on Wikipedia, WebMD and Mayo Clinic websites, this study represents a response to a significant gap. Findings are anticipated to assist consumers in gathering high quality and complete health information. The study will also benefit care providers in understanding the strengths and limitations of the information available online.

Methods

Development and selection of statements as basis of tool

We developed a specific tool to address the scope, completeness, credibility, and readability criteria of the health statements. Although several validated measures of health literacy skills exist such as the Test of Functional Health Literacy in Adults (TOFHLA) (Parker, Baker, Williams, & Nurss, 1995) in Adults, the Rapid Estimate of Adult Literacy (REALM) (Davis, Long, & Jackson, 1993) and the Newest Vital Sign (NVS) (Weiss et al., 2005) they do have their limitations.

The TOFHLA is lengthy to administer (22 minutes or more for the full version and up to 10 minutes for the abridged version). THE REALM is quick to administer (3 minutes) but does not examine readability (Rowlands et al., 2013) or scope. The NSV does not capture the four criteria that are important in evaluating health, medical and nutritional information online.

According to the Pew Internet and American Life project that the most specific researched queries are related to medical, health, nutrition and general health topics (Pew Internet: Health, 2013). We developed 111 health statements (from these four topics areas) and through an iterative and collaborative process with the research team, agreed to remove open ended, misleading and ambiguous statements. Based on the abovementioned four topics, 92 statements were remained. These 92 statements represented nine subcategories: (a) general health (20 statements), (b) food or nutritional information (18 statements), (c) specific diseases (16 statements), (d) mental health (13 statements), (e) weight loss and management (11 statements), (f) pregnancy and childbirth (6 statements), (f) specific medical treatments or procedures (4 statements), (g) cardiovascular disease (2 statements), and (h) medical tests (2 statements). The nine subcategories were created based on the findings of the Pew Internet and American Life project in which the most commonly searched subjects by health consumers were identified (Pew Internet & American Life Project, 2014). Based on these search subjects, our statements were classified into the nine categories.

Scope, completeness, credibility, and readability

Scope was recorded as a binary outcome (yes or no). Each rater received five hours of training from the principal investigator. Health statements that scored a 'yes' for the following statement, 'Is the website able to answer the question?' received a score for completeness. A 3-point scale was used to determine completeness, with 3 being 'fully complete', 2 being 'partially complete', and 1 being 'very incomplete'.

Credibility was determined based on whether or not references supporting the statement were included on the webpage. References were further classified according to the following categories: blogs, university or academic websites, textbooks, peer-review journals, non-profit/community organization (e.g., cancer society, heart and stroke, etc.) and other. The frequency of updates to individual webpages was also documented.

According to Calderon et al., (2006) the Flesch Reading Ease Score (FRES) and Flesch-Kincaid Grade Level (FKGL) formulas are commonly used to assess readability (Calderón, Morales, Liu, & Hays, 2006). The FRES rates text on a 100-point scale. The formula is 206.835 – (1.015 x ASL) – (84.6 x ASW) where ASL = average sentence length (no. words/no. sentences) and ASW = average syllables per word (Microsoft Word, 2013). Such formulas generally involve counting the number of syllables per word and the number of words per sentence. The FRES is one of the most widely and validated measures of readability and is widely accepted by the insurance industry for evaluating documents intended to be read by consumers (Graber et al., 2002). A lower FRES (e.g., 30-49) indicates that the material is more difficult to read. A score of 60 or greater which corresponds with a secondary school reading level is considered to be a minimally acceptable level for consumer-oriented information (Bernstam, Shelton, Walji, & Meric-Bernstam, 2005).

The Flesch-Kincaid Grade Level score rates text according to an American school grade level and measures reading level. For instance, a score of 7.0 indicates a seventh grade understanding. In this study, the FKGL formula was also used to determine readability. The FKGL is appealing because of its use in commercial word processing software(Microsoft Word 2007) (Estrada, Hryniewicz, Higgs, Collins, & Byrd, 2000). The formula as used in the software is (0.39 × ASL) + (11.8 × ASW) − 15.59, where ASL is the average sentence length (number of words divided by number of sentences) and ASW is the average syllables per word (number of syllables divided by number of words) (Microsoft Word, 2013; Strawbridge, 2008).

Development of questionnaire tool

A four-page English language questionnaire was developed that included the criteria described above regarding scope, completeness, credibility, and readability of the health information assessed in relation to the 92 health statements. The questionnaire was based on existing literature about health, nutrition, and medicine (Clauson, Polen, Boulos, & Dzenowagis, 2008; Giles, 2005; Haigh, 2011; Kupferberg & Protus, 2011; Lavsa, Corman, Culley, & Pummer, 2011; Leithner et al., 2010; Reavley et al., 2012; Volsky, Baldassari, Mushti, & Derkay, 2012). Furthermore, it was specifically development to examine the scope, completeness, credibility, and readability of the health information of health, nutritional and medical information. We pilot tested the tool among a smaller set (30 health statements) to refine the reliability and validity of the measure.

The questionnaire included three sections: i) scope & completeness (e.g., does the article answer the question completely?); ii) readability (Flesh Reading Ease Score and the Flesch-Kincaid Grade Level); and iii) credibility (sources used to support health statement). The reviewers noted any further information related to the statements on the website (Wikipedia, WebMD, and Mayo Clinic) that were not captured by the questionnaire in open text boxes.

Training of raters

Each rater participated in a five-hour training session. The training consisted of data entry, using systematic search keywords, search strategies, and the use of an electronic log document for each health statement. Each rater was instructed to evaluate the scope, completeness, credibility, and readability ten practice health statements prior to starting the formal standardized searches. After training, the raters entered each of the 92 statements into the three health repository websites using a standardized search protocol (see below), an electronic log of their search terms, and content identified on each website that addressed the scope, readability and credibility of the health statements.

Standardized searches

Trained raters entered the exact same 92 health statements (e.g., what is the role of folic acid in healthy fetal development?) in the search fields of Wikipedia, WebMD and the Mayo Clinic websites. All links on the first electronic page for each search engine were then counted, classified and visited. Raters then followed the relevant links until they were able to obtain the pertinent information regarding scope, completeness, credibility, and readability of the health information of each health statements. If the rater was unable to answer the question after visiting all the links

on the first page of the search results (typically after visiting ten web links), the search was discontinued. For all three website searches, the search box on the main page was used to conduct the queries.

Each rater used an electronic log file that included webpages examined, a list of keywords, and evidence used to evaluate the statements. Subsequently, each rater used this approach for a total of 276 entries from February to April, 2013. Each rater made notes in their electronic file as well as the database file regarding points of disagreements, points of clarification or questions. These issues were resolved through discussion and in a collective fashion with the research team.

Quality assurance

Once data were collected and entered in STATA version 11.0, (StataCorp, 2009) several data validation checks were performed for logical consis-ency. Outliers, missing data values, and suspicious entries were also identified.

The inter-rater agreement was evaluated using Cohen's Kappa and all of the statements were coded by the primary rater (BN) (J. Cohen, 1960). A second rater (RK) was provided a random sample of 25% of the statements for the three websites. The second rater coded findings independent of the first rater for 23 statements. Inter-rater reliability for the raters was high for completeness, readability, and credibility. Cohen's Kappa was between 0.95-1.00, indicating very close agreement between the two coders (Carletta, 1996).

Data analyses

The units of analysis were the links (specific uniform resource locator) from the three websites, the standardized rating form used for evaluation of quality, and content used for the determination of grade reading level. Descriptive and summary measures (percentages, frequencies, and cross-tabulations) were calculated for all variables (e.g., Fisher's exact two-tailed test was used to investigate scope, and completeness scores between databases. Fleiss' Kappa was used to measure reliability of agreement between multiple raters (Fleiss, 1971). Interclass correlation coefficient (ICC) (Deyo, Diehr, & Patrick, 1991) was used as a measure for describing the reliability and validity of the data and tested the degree of association, slope, and intercept variance expected from replicated measures. Comparisons were based on the Kruskal-Wallis test, which examined means and assessed the relationship between the two methods of determining readability.

Integrated model of health literacy

We used the model proposed by (Sørensen et al., 2012) that combines the qualities of a conceptual model outlining the main dimensions of health literacy, and of a logical model showing the proximal and distal factors which impact on healthy and their relationship to health outcomes (Sørensen et al., 2012). The model depicts the competencies related to the process of access, understanding, appraising and applying health-related information (Sørensen et al., 2012). Access refers to the ability to seek, find and obtain health information. Understanding refers to the ability to comprehend the health information that is accessed; appraising denotes the health information that has been access; and applying demonstrates the ability to communicate and use the information to make a decision to maintain and improve health. Each of the four dimensions represents a crucial dimension of health literacy, requires specific cognitive qualities and depends on the quality of the information provided (Sørensen et al., 2012). We used the integrated model of health literacy to assist in examining the scope, completeness, credibility, and readability of health, medical, and nutritional information on the Wikipedia, WebMD and Mayo Clinic websites.

Results

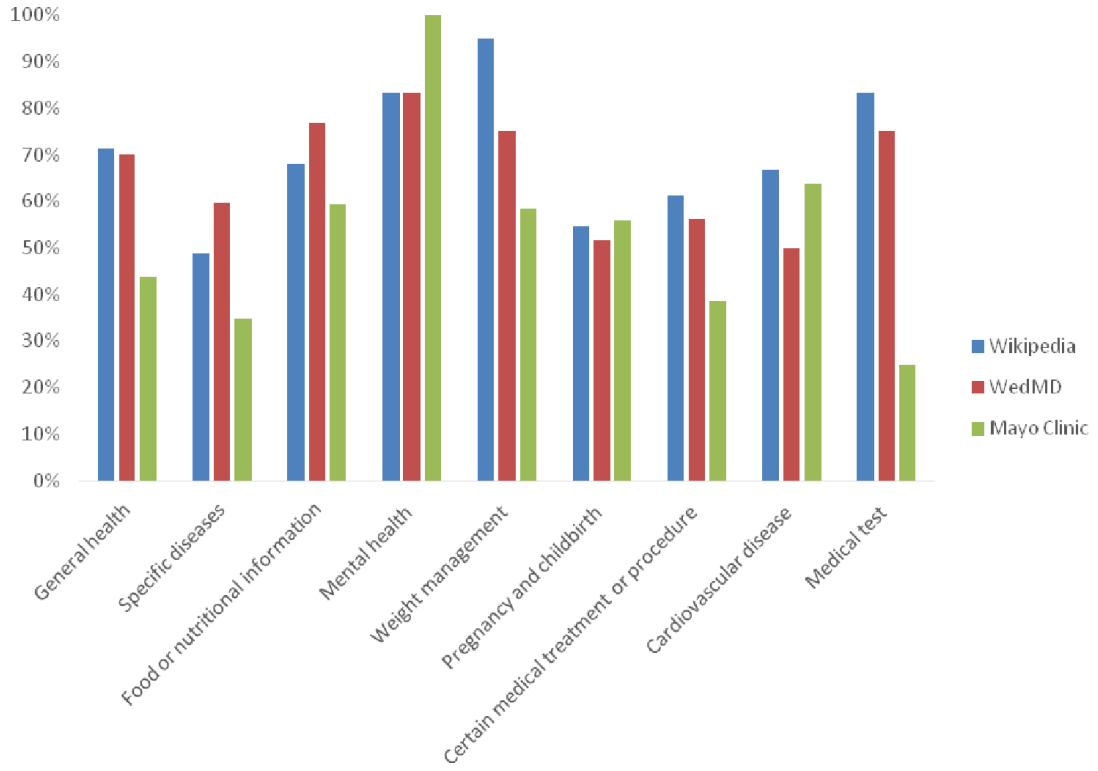
Scope

Wikipedia provided answers to 91.3% of the 92 general health statements; WebMD (89.1%) and the Mayo Clinic (81.5%) followed respectively. Fleiss' Kappa statistic (strength of agreement) was 0.85 and considered excellent according to the Fleiss' Kappa Benchmark Scale (Fleiss, 1981).

Completeness

Overall, 67.7% of Wikipedia pages provided complete answers, followed by WebMD (66.1%), and the Mayo Clinic (50.2%). Each website was also evaluated according to the defined health statement categories. For example, the Mayo Clinic provided the most complete answers for 100% of the mental health category statements, followed by WebMD (83.0%), and Wikipedia (83.0%). Completeness scores (%) for Wikipedia, WebMD, and Mayo Clinic websites are shown in Figure 1. Fleiss' Kappa statistic was 0.53 and considered intermediate to good according to the Fleiss' Kappa Benchmark Scale (Fleiss, 1981).

Figure 1 *Completeness* Scores (%) by Health Category*



Reading Grade Levels

The mean and standard deviation FRES for Wikipedia was 26.7 ± 14.1. For WebMD, the values were 43.9 ± 19.4 while the values for the Mayo Clinic website were 40.9 ± 19.4. The range in means across the three websites is described as "very difficult" (Wikipedia) (e.g., analogous to reading the *New England Journal of Medicine*) to "difficult" (Mayo Clinic) (e.g., analogous to reading novels such as the Henry James novel, *The Ambassadors*) (Roberts, Fletcher, & Fletcher, 1994; White, Jones, Felton, & Pool, 1996). FRES was also examined by health category (Table 1). More difficult passages included numerous multisyllabic words, long sentences, and difficult sentence syntax and structures. The FRES scores were significantly higher for Wikipedia compared to WebMD and the Mayo Clinic websites ($p<0.001$). The mean FKGL grade levels ranged from 10th to the 14th (college level) grade. The mean and standard deviation for Wikipedia was 14.5 ± 5.0; 10.2 ± 3.9 for WebMD and 10.2 ± 4.4 for the Mayo Clinic website. The FKGL scores were significantly higher for Wikipedia compared with those for WebMD and the Mayo Clinic websites ($p<0.001$).

For the FRES, a strong correlation is found between the three raters across 92 questions from the three sites (Wikipedia, WebMD, and Mayo Clinic) (ICC (2,3) = 0.73). Siimilarly, for the FKGL, a moderate correlation is reported average between the three raters (ICC (2,3) = 0.57).

Table 1 *Mean and Standard Deviation Flesch Reading Ease Scores (FRES) by Health Category*

	Wikipedia M, SD(n)	WebMD M, SD(n)	Mayo Clinic M, SD(n)
General health	28.5,14.3(59)	45.2, 19.8(60)	42.7, 19.4(57)
Specific diseases	25.7, 15.1(49)	45.8, 19.0(49)	44.7, 17.5(43)
Food or nutritional information	30.4, 10.1(47)	42.5, 20.2(40)	35.5, 22.4(42)
Mental health	23.3, 7.7(6)	47.6, 25.3(6)	40.7, 20.4(6)
Weight management	20.2, 13.4(38)	36.7, 18.1(36)	34.4, 19.6(36)
Pregnancy and childbirth	28.1, 15.1(33)	54.2, 16.1(31)	46.1, 15.5(26)
Certain medical treatment or procedure	322, 17.5(18)	43.5, 19.2(16)	51.3, 20.6(13)
Cardiovascular disease	18.1, 9.5(12)	39.0, 16.8(12)	36.3, 14.2(12)
Medical test	24.0, 10.5(6)	32.5, 20.1(4)	35.8, 9.0(4)

Credibility

Sources supporting the statement were present for 96.0% of webpages on the Mayo Clinic site, 95.1% of webpages for Wikipedia, and 94.9% of webpages for WebMD. There were no statistically significance differences between the three groups. Peer reviewed articles were the most frequently cited reference on all three websites. Table 2 provides information on the credibility of the three websites.

Table 2 *Sources supporting health statements for Mayo Clinic, WebMD, and Wikipedia*

	Wikipedia n | %	WebMD n | %	Mayo Clinic n | %
Sources are presented in the article			
Yes	253 | 95.1	242 | 96.0	224 | 94.9
No	13 | 4.9	10 | 4.0	12 | 5.1
Total	266 | 100	252| 100	236 | 100
Types of references used			
Peer reviewed journals	125 | 56.8	67 | 31.9	80 | 40.2
Non-profit/community organization	16 | 7.3	30 | 14.3	16 | 8.0
University or academic websites	24 | 10.9	49 | 23.3	31 | 15.6
Textbooks	23 | 10.5	4 | 1.9	11 | 5.5
Blogs	0 | 0	3 | 1.4	2 | 1.0
Combination of references	10 |4.5	26 | 12.4	31 | 15.6
Other sources	22 | 10.0	31 | 14.8	28 | 14.1
Total	139| 100	210 | 100	220 | 100
Revision dates provided			
6 months or less	252 | 94.0	59 |23.8	47 | 20.0
1 year or less	10 | 3.7	102 |41.1	87 | 37.0
>1 year	6 | 2.3	87 | 35.1	101| 43.0
Total	268| 100	248 | 100	235 | 100

Discussions

The purpose of the study was to examine the scope, completeness, credibility, and readability of health, medical, and nutritional information found on Wikipedia, WebMD and Mayo Clinic websites. Answers from the Mayo Clinic website were less complete compared with those found at WebMD and Wikipedia. In terms of scope, Wikipedia provided the greatest number of answers to the 92 general health statements across nine categories, followed by WebMD, and the Mayo Clinic websites. Wikipedia also provided the most complete answers across the nine health categories.

Based on these findings, Wikipedia appears to be the most useful source of health information of the three sites. However, because of the breadth of information found on Wikipedia, the site may not hold broad appeal to general readers, and there may be potential for misinterpretation and misunderstanding of health information. Furthermore, Wikipedia does not provide practical guidance or recommendations when compared to the WedMD and Mayo Clinic websites. As a general observation, consumers who rely on incomplete and inaccurate health information risk not learning about important safety precautions such as contraindications and drug interactions.

We also examined readability scores between

our three raters across reading grade levels. The mean ratings for the FRES were "difficult" for WebMD and Mayo Clinic websites and "very difficult" for Wikipedia. Given that nearly a third of American patients have basic or below basic health literacy (U.S. Department of Health & Human Services, 2008), these findings emphasize the importance of communicating health information in language that is easy to comprehend.

Healthy People provides science-based, 10-year national objectives for improving the health of Americans (HealthyPeople.gov, 2013). The Health People 2020 has an objective is to improve health literacy among the general population. Therefore, this study provides timely evidence of the importance of evaluating health information online and promoting health literacy.

With improved health literacy, health consumers must also become empowered to use online sources of health information to make better and more informed decisions about their own health (Hay, 2009). Wikipedia addresses reading difficulty by including an alternate simple English version of its articles aimed at lower reading levels and non-native readers. The investigators recommend a collaboration involving scientific writers, clinicians, and research scientists to create an online database written at the user's reading level with the most up to date information. Such an undertaking would involve collaboration among multiple institutions, with final editing by designated authors. In a sense, this model is a kind of Wikipedia resource written by scientific authors and vetted by appropriate medical expertise. Currently, websites like UpToDate (http://www.uptodate.com) provide such a service for healthcare professionals, but no such health information portal exists for the general public.

For Wikipedia, article authorship is unrestricted and, on the whole, articles are unedited. This makes the Internet a dynamic and ripe medium for information exchange. While free expression on the Internet facilitates communication, it can also be at odds with the ideals of scope, and completeness which medical and scientific literature hold sacred (Clauson et al., 2008; Volsky et al., 2012). Given the rapidly changing nature of online health information, users are strongly encouraged to educate themselves and critically assess all information. It is possible that Wikipedia's authors access the other two sites, thus generating a positive bias. Such an idea merits its own separate investigation.

While interesting, the results must be taken with caution. First, the study focused on English-language resources within three popular health information websites. Based on this, the external validity of the study beyond specific audiences and nations (mainly the United States and Canada) is threatened. Moreover, only three websites were assessed for a finite number of categories, and it is possible that an evaluation of different health websites may render different results. As well, we used standardized searches and raters to focus on the most relevant health categories as identified by recent survey (Pew Internet & American Life Project, 2014), therefore, we may have missed some health topics that were not identified by the survey results.

Inherently, with every tool, its psychometric properties need to be rigorously evaluated. However, health literacy is not consistently measured, making it difficult to interpret and compare at individual and population levels (Jordan, Osborne, & Buchbinder, 2011). As such, we used a tool that examined the scope, completeness, credibility, and readability of health, medical, and nutritional information found on these three websites. Therefore, empirical evidence demonstrating validity and reliability of existing indices is required, and more comprehensive health literacy instruments need to be developed (Jordan et al., 2011).

Conclusions

This study is a first effort at comparing the scope, completeness, credibility, and readability of health, medical, and nutritional information on Wikipedia, WebMD and Mayo Clinic websites. These are websites geared to the general public. The varying levels of scope, completeness, and readability found in the study reinforce the need for such studies and for expanding them to include web pages in other languages. Only in this way can researchers and health providers acquire a more inclusive understanding of where Americans seek their health information. Because the Internet is an important source of medical and general health information and changes rapidly (Bennett, Casebeer, Kristofco, & Strasser, 2004) such information must be regularly evaluated and assessed. Future studies that include other repositories of health information serving health consumers are recommended. The importance of the quality of health care information available on the Internet for patients is a very important topic and of great interest for health consumers and health care professionals. Expansion of the evaluation criteria for in-depth analysis of supporting evidence, biases, and revision patterns are likewise suggested.

Conflict of interest

The authors of this paper do not have any commercial associations that might pose a conflict of interest in connection with this manuscript.

References

Alexa - The Web Information Company. (2013). Wikipedia.org. Retrieved April 23, 2013, from http://www.alexa.com/siteinfo/wikipedia.org

Bennett, N. L., Casebeer, L. L., Kristofco, R. E., & Strasser, S. M. (2004). Physicians' Internet information-seeking behaviors. Journal of Continuing Education in the *Health* Professions*, 24*(1), 31-38.

Berland, G. K., Elliott, M. N., Morales, L. S., Algazy, J. I., Kravitz, R. L., Broder, M. S., , Kanouse, D..E., Munoz, J.A., Lara, M., Watkins, K.E., Yang, H., McGlynn, E.A.. (2001). Health Information on the Internet: Accessibility, Quality, and Readability in English and Spanish. *The Journal of the American Medical Association 285*(20), 2612-2621.

Bernstam, E. V., Shelton, D. M., Walji, M., & Meric-Bernstam, F. (2005). Instruments to assess the quality of health information on the World Wide Web: what can our patients actually use? *International journal of medical informatics, 74*(1), 13-20.

Calderón, J. L., Morales, L. S., Liu, H., & Hays, R. D. (2006). Variation in the readability of items within surveys. *American Journal of Medical Quality, 21*(1), 49-56.

Carletta, J. (1996). Assessing agreement on classification tasks: the kappa statistic. *Computational Linguistics, 22*, 249-254.

Clauson, K. A., Polen, H. H., Boulos, M. N. K., & Dzenowagis, J. H. (2008). Scope, completeness, and accuracy of drug information in Wikipedia. *Ann Pharmacothe* Annals of *Pharmacotherapy , 42*(12), 1814-1821.

Cohen, J. (1960). A coefficient of agreement for nominal scales.Educational and Psychological Measurement *20*, 37-46.

Cohen, R., Elhadad, M., & Birk, O. (2013). Analysis of Free Online Physician Advice Services. *PloS one, 8*(3), e59963.

Couper, M. P., Singer, E., Levin, C. A., Fowler, F. J., Fagerlin, A., & Zikmund-Fisher, B. J. (2010). Use of the Internet and ratings of information sources for medical decisions: results from the DECISIONS survey. *Medical Decision Making, 30*(5 suppl), 106S-114S.

Davis, T., Long, S., & Jackson, R. (1993). Rapid estimate of adult literacy in medicine: a shortened screening instrument. Family Medicine *, 25*, 39-35.

Deyo, R. A., Diehr, P., & Patrick, D. L. (1991). Reproducibility and responsiveness of health status measures statistics and strategies for evaluation. *Controlled clinical trials, 12*(4), S142-S158.

Estrada, C. A., Hryniewicz, M. M., Higgs, V. B., Collins, C., & Byrd, J. C. (2000). Anticoagulant patient information material is written at high readability levels. *Stroke, 31*(12), 2966-2970.

Fleiss, J. L. (1971). Measuring nominal scale agreement among many raters. *Psychological bulletin, 76*(5), 378.

Fleiss, J. L. (1981). *Statistical Methods for Rates and Proportions*. New York: John Wiley and Sons:.

Giles, J. (2005). Internet encyclopaedias go head to head. *Nature, 438*(7070), 900-901.

Graber, M. A., D Alessandro, D. M., & Johnson-West, J. (2002). Reading level of privacy policies on internet health web sites. *Journal of Family Practice, 51*(7), 642-642.

Haigh, C. A. (2011). Wikipedia as an evidence source for nursing and healthcare students. *Nurse Educ Today, 31*(2), 135-139.

Hay, L. (2009). In practice. *Perspect Public Health Perspectives in Public Health, 129*(4), 156.

HealthyPeople.gov. (2013, April 23, 2013). Health communication and health information technology: HC/HIT-1 (Developmental) Improve the health literacy of the population. Retrieved May 17, 2014, from http://www.healthypeople.gov/2020/topicsobjectives2020/objectiveslist.aspx?topicid=18

Helft, P. R. (2008). A new age for cancer information seeking: are we better off now? *Journal of general internal medicine, 23*(3), 350-352.

Jordan, J. E., Osborne, R. H., & Buchbinder, R. (2011). Critical appraisal of health literacy indices revealed variable underlying constructs, narrow content and psychometric weaknesses. *Journal of clinical epidemiology, 64*(4), 366-379.

Kupferberg, N., & Protus, B. M. (2011). Accuracy and completeness of drug information in Wikipedia: an assessment. *Journal of the Medical Library Association 99*(4), 310.

Lavsa, S. M., Corman, S. L., Culley, C. M., & Pummer, T. L. (2011). Reliability of Wikipedia as a medication information source for pharmacy students. *Currents in Pharmacy Teaching and Learning, 3*(2), 154-158.

Leithner, A., Maurer-Ertl, W., Glehr, M., Friesenbichler, J., Leithner, K., & Windhager, R. (2010). Wikipedia and osteosarcoma: a trustworthy patients' information? *Journal of the American Medical Informatics Association, 17*(4), 373-374.

Mayo Clinic. (2013). *Health Information*. Retrieved April 23, 2013, from http://www.mayoclinic.com/health-information/

McMullan, M. (2006). Patients using the Internet to obtain health information: how this affects the patient–health professional relationship. *Patient education and counseling, 63*(1), 24-28.

Microsoft Word. (2013). Test your document's readability. Retrieved April 23, 2013, from http://office.microsoft.com/en-ca/word-help/test-your-document-s-readability-HP010148506.aspx

National Patient Safety Foundation. (2011). Health Literacy: Statistics at-a-glance. Retrieved April 23, 2013, from http://www.npsf.org/wp-content/uploads/2011/12/AskMe3_Stats_English.pdf

O'Reilly, K. (2012). American Medical News. The ABCs of health literacy. Retrieved April 23, 2013, from http://www.amednews.com/article/20120319/profession/303199949/4/

Parker, R. M., Baker, D. W., Williams, M. V., & Nurss, J. R. (1995). The test of functional health literacy in adults. *Journal of general internal medicine, 10*(10), 537-541.

Pew Internet & American Life Project. (2014). Information Triage. Retrieved July 1, 2014, from http://www.pewinternet.org/Reports/2013/Health-online/Part-One/Section-5.aspx

Pew Internet: Health. (2013, April 23, 2013). Pew Internet: Health. Retrieved July 2, 2014, from http://www.pewinternet.org/Commentary/2011/November/Pew-Internet-Health.aspx

Reavley, N. J., Mackinnon, A. J., Morgan, A. J., Alvarez-Jimenez, M., Hetrick, S. E., Killackey, E., Nelson, B., Purcell, R.., Yap, M.B., Jorm, A. F. (2012). Quality of information sources about mental disorders: a comparison of Wikipedia with centrally controlled web and printed sources. *Psychological medicine, 42*(8), 1753.

Roberts, J. C., Fletcher, R. H., & Fletcher, S. W. (1994). Effects of peer review and editing on the readability of articles published in Annals of Internal Medicine*., The Journal of the American Medical Association 272*(2), 119-121.

Rowlands, G., Khazaezadeh, N., Oteng-Ntim, E., Seed, P., Barr, S., & Weiss, B. D. (2013). Development and validation of a measure of health literacy in the UK: the newest vital sign. *BMC public health13*(1), 116.

Sørensen, K., Van den Broucke, S., Fullam, J., Doyle, G., Pelikan, J., Slonska, Z., & Brand, H. (2012). Health literacy and public health: a systematic review and integration of definitions and models. *BMC public health 12*(1), 80.
StataCorp. (2009). Stata statistical software: release 11.0. College Station, TX: Stata Corporation.
Strawbridge, M. (2008). *Microsoft Office Word 2007 Essential Reference for Power Users*. Cambridge, UK: Software Reference Ltd.
U.S. Department of Health & Human Services. (2008). America's Health Literacy: Why We Need Accessible Health Information. An Issue Brief Retrieved July 4, 2014, from http://www.health.gov/communication/literacy/issuebrief/
Volsky, P. G., Baldassari, C. M., Mushti, S., & Derkay, C. S. (2012). Quality of Internet information in pediatric otolaryngology: A comparison of three most referenced websites. *International Journal of Pediatric Otorhinolaryngology 76*(9), 1312-1316.
Weiss, B. D., Mays, M. Z., Martz, W., Castro, K. M., DeWalt, D. A., Pignone, M. P.,, Mockbee, J., Hale, F. A. (2005). Quick assessment of literacy in primary care: the newest vital sign. *The Annals of Family Medicine, 3*(6), 514-522.
White, L. J., Jones, J. S., Felton, C. W., & Pool, L. C. (1996). Informed consent for medical research: common discrepancies and readability. *Acad Emerg Med 3*(8), 745-750.
Wikipedia. (2014). Wikipedia:About. Retrieved July 15, 2014, from https://en.wikipedia.org/wiki/Wikipedia:About