Accuracy of Ebola Information in a Knowledge Exchange Social Website (KESW)

Fiona Gorman1*, Desa Yadegarians1, Taufa Islam1, Simon Tongco2, Esther Johnston3, Elma Estrada2, Nicholas Gorman4

1California State University, Long Beach, USA
2California State University, Fullerton, USA
3The Wright Center National Family Medicine Residency at Health Point, Auburn, USA
4Keck Graduate Institute, Claremont, USA
Email: *Fiona.gorman@csulb.edu

Abstract

Background: Misinformation on interactive Knowledge Exchange Social Websites (KESWs) is concerning since it can influence Internet users’ health behaviors, especially during an infectious disease outbreak. Objective: The present study seeks to examine the accuracy and characteristics of health information posted to a Knowledge Exchange Social Website (KESW). Methods: A sample of 204 answers to Ebola questions were extracted and rated for accuracy. Multiple logistic regression modeling was used to examine whether answer characteristics (best answer, professional background, statistical information, source disclosed, link, and word count) predicted accuracy. Results: Overall, only 27.0% of the posted answers were rated as “accurate”. Accuracy varied across question topics with between 11.8% - 45.5% of answers being rated as accurate. When Yahoo Answers’ “best answers” were examined, the overall accuracy was substantially higher, with 80.0% of “best answers” being rated as accurate compared to 16.0% of all other answers. Conclusion: There is need for tools to help Internet users navigate health information posted on these dynamic user-generated knowledge exchange social websites.

Keywords

Online Health Information, Ebola, Infectious Diseases, Accuracy, Knowledge Exchange Social Websites

1. Introduction

The popularity of the Internet as a discreet, readily available source of health in-
formation is evidenced by data showing that up to 75% of adults in the U.S. report having used a search engine to look up health information [1], with over half of adults (56.8%) having reported seeking health information online in the past month [2]. Daily, approximately three times as many people search for health information online compared to consulting physicians at office visits [3].

Despite the popularity of online health resources, past research has shown that both the quality [4] [5] and readability level [4] of health-related websites varies widely, with incomplete information and inaccuracies compromising the information available to readers [6].

The presence of poor, incomplete, or misleading information is troubling given the low levels of electronic health literacy (eHealth Literacy) reported among Internet users [6] [7]. Most Internet users studied reported that they knew how to find health information, but their confidence in distinguishing high quality from low quality resources online was significantly lower [8] [9]. Even among students of health science, low levels of eHealth Literacy have been documented, with few knowing about free, credible health databases [10].

The impact of misinformation online is at least twofold. First, the mere presence of such information can influence Internet users’ search and browsing habits. For instance, confirmation biases, in which people tend to seek out information that confirms their preexisting beliefs [11], have been observed to operate in online searches. In the presence of conflicting information on a health topic, individuals actively and preferentially access information that reinforces their beliefs while avoiding information that challenges their beliefs [12]. Similarly, research has shown that internet users’ search strategies are systematically biased towards examining only the top search results from search engines and following links related to more serious health conditions when trying to self-diagnose [13]. Despite these biases, Internet users tend to believe the information they find online is accurate and trustworthy, regardless of the actual accuracy of the information [14].

Second, health information has also been shown to actively shape various users’ health attitudes and behaviors [1] [15] [16] [17] [18] [19]. Research shows that online health information can influence when or if users contact health professionals [20], participate in health screening programs [21], engage in complementary and alternative medicine [22], or adhere to physician recommendations [21] [23] [24].

Previous research has examined the quality and accuracy of health information posted to professional, static websites [25] [26] and online supplement retailers [27]. While poor, incomplete, or misleading information is present on many of these websites, researchers have made efforts to address this through the development of tools for assessing the veracity of such pages’ content [28] and noted several characteristics of higher quality sources such as the presence of disclaimers, availability of references, and authorship disclosure [29] as well as the length of information and frequent external links [30].
Ascertaining the veracity of webpages’ health information becomes more challenging on interactive websites where Internet users play a role in generating the web pages’ content. Whereas the overall quality and accuracy of a static website’s content, which is developed by a single person or single team, can often be assessed in a single pass, such assessments become more challenging with user-generated content, where high-quality health information can be presented alongside incomplete information, emotionally powerful personal anecdotes, misinformation attributable to users’ misperceptions, and even intentionally misleading misinformation.

While research has examined the overall quality of information on some interactive health-focused websites such as the message boards used in online support groups [31], relatively little is known about less health-specific websites, such as knowledge exchange social websites (KESWs). Knowledge Exchange Social Websites differ from support group message boards in that they are not explicitly focused on a single topic. Whereas support groups develop communities defined by a shared interest in a single, specific health condition, KESWs, which provide a forum for users to post questions on any topic anonymously and to respond to other users’ questions anonymously, attract a much broader, diverse audience of users.

Accurate and timely online information is particularly important during an outbreak of a (re)emerging infectious disease. Slow dissemination of information through official channels and confusing or conflicting messages in the media generate high levels of panic in the general public and drive them to seek answers on the internet [32] [33] [34]. The current study focuses on Ebola virus disease (EVD), as the response to the 2014 outbreak in West Africa was impacted by the presence of misinformation and highlighted the effects of such information on outbreak containment, support of proper quarantine procedures, and social stigmatization of patients [35] [36]. While the 2014 outbreak never made significant inroads into the United States, research has highlighted several characteristics of the US populace that could exacerbate the spread of an infectious disease during a global pandemic. Specifically, knowledge and utilization of official channels for government health communication remain low [36]. At the same time, in the case of Ebola, overall knowledge about the disease is low [36] [37] while generalized mistrust and conspiracy beliefs related to the medical industry [35] are prevalent. Under conditions in which Internet users are underutilizing official health communication channels, harboring mistrust towards the medical establishment, and carrying factual inaccuracies about a disease, KESWs, with their anonymity, pose a risk for fueling the spread of misinformation.

The present study seeks to address some of the knowledge gap on the accuracy of health information posted to KESWs by examining the types of Ebola questions being posted on a popular KESW and rating the accuracy of the anonymous users’ answers to these questions. In addition, the relationship between an-
swer characteristics, such as inclusion of links to references, and answers’ accuracy was examined in order to determine whether answer characteristics could be used to identify higher quality answers.

2. Methods

2.1. Data Collection

The decision was made to focus the study on a single KESW. Of the KESWs reviewed, Yahoo Answers was selected due to the interface’s ease of searching and retrieving questions and answers as well as for its reach; in 2016 Yahoo was ranked as the third most popular multi-platform web property in the United States with 206 million unique visitors in a single month [https://www.statista.com/statistics/271412/most-visited-us-web-properties-based-on-number-of-visitors/]

On March 25, 2015, a total of 23 posts with the keyword “ebola” were extracted from Yahoo Answers for analysis (see Figure 1). Upon initial review 5 posts were excluded as they asked subjective questions whose answers could not be rated for accuracy (see Table 1’s excluded category for an example question),

![Flow chart of question/answer inclusion and exclusion.](image)
resulting in a dataset of 18 posts. Upon further review, several of the 18 posts contained multiple questions. Each question within the posts was examined independently, yielding a total of 35 distinct questions about Ebola. A total of 204 answers were offered to these 35 questions. Each question had between 2 to 11 answers offered, with the average number of answers posted per question being 5.83 (SD = 3.24).

In addition to questions and answers, six accompanying data points were extracted from each answer:

1) Best Answer: Since March 2014, the person who posted their question(s) on Yahoo Answers gets to mark one of the answers provided as the Best Answer. All sets of answers had a Best Answer marked.

2) Professional Background: This variable captured whether or not each answerer indicated that their answer was based on their professional background in the health sciences (ex: answerer indicated that they were a nurse with 10 years of experience with infectious diseases).

3) Statistical Information: This variable captured whether or not each answer included the use of statistics.

4) Source Disclosure: This variable captured whether or not each answer contained a disclosure that the information presented came from an external source, as it was discovered that many answers contained unmodified copied and pasted content from other websites.

5) Link: This variable captured whether the answer contained a link to an external website for additional information.

6) Word Count: A count of the words used in each answer.

2.2. Answer Accuracy

In order to evaluate the accuracy of each posted answer, answers were coded into one of five categories:

1) Accurate: Accurate answers contained no factual errors and addressed the question that was asked.

2) Inaccurate: Inaccurate answers contained one or more factual errors. Note that, given the severe consequences of misinformation on infectious diseases, it was decided to rate answers as inaccurate even if the answer contained accurate information as well as inaccurate information.

3) Subjective: Subjective answers included any response whose accuracy could not be rated, such as statements of opinion.

4) Unanswered: Unanswered answers represented responses that did not address the question that was asked.

5) Trolling: Upon working with the data, it became clear that a fifth category was needed in order to capture responses that not only didn’t answer the question asked, but which also took on the characteristics of online trolling, which Merriam-Webster defines as “to antagonize (others) online by deliberately posting inflammatory, irrelevant, or offensive comments or other disruptive content”
The accuracy of all answers was assessed independently by two of the authors. The authors then examined each other’s ratings and discussed the answers they disagreed upon. A physician was available as the tiebreaker in case the authors could not agree upon an answer’s accuracy rating after discussion, though all disagreements were resolved with discussion between the authors without need for the physician’s intervention.

3. Data Analysis

All data were analyzed with SPSS version 24.0 (IBM, 2016) [38]. A thematic analysis was conducted in order to establish a codebook of the types of questions being asked about Ebola on Yahoo Answers [39]. In the first stage, two of the authors read through the entirety of the set of questions in order to familiarize themselves with the data. Following the read-through both readers independently developed a set of emergent themes to organize the types of questions asked. These emergent themes were then shared with the full research team who helped to reconcile differences in the two authors’ coding schemes and arrive at a final coding scheme.

Simple descriptive statistics (frequency and valid percent) and histograms were employed to examine the types of Ebola questions being asked, the accuracy of answers to these questions, and the role of answers voted “best answer” by the KESW user who posted each question.

Multiple logistic regression modeling was used to examine whether answer characteristics (best answer, professional background, statistical information, source disclosed, link, and word count) predict accuracy (re-coded to a dichotomous accurate vs. inaccurate). Answers that fundamentally failed to address the question asked (i.e. subjective, trolling, or unanswered) were excluded from the logistic regression model, as readers looking for an answer to a health question could reasonably be expected to disregard these answers. As there were no a priori predictions regarding which variables would emerge as significant predictors of answers’ accuracy, five of the six predictors were force entered into the final logistic regression model. The sixth predictor, professional background, was ultimately removed from the model, as only three answers came from respondents citing a professional background, which precluded meaningful analysis of this variable.

4 Results

4.1. Types of Ebola Questions Asked

A total of seven themes were identified during the thematic analysis of types of Ebola questions posted to Yahoo Answers. Table 1 presents each theme along with a representative example question drawn from the dataset.

The topics of Yahoo Answers visitors’ questions showed significant heterogeneity, with each of the question categories capturing only between 4.9% -
**Table 1.** Types of Ebola Questions Posted to a KESW (n = 209).

<table>
<thead>
<tr>
<th>Question Topic</th>
<th>n</th>
<th>Example Question:</th>
</tr>
</thead>
<tbody>
<tr>
<td>Biology</td>
<td>56</td>
<td>What can you tell me about the structure of the Ebola virus?</td>
</tr>
<tr>
<td>Causes</td>
<td>17</td>
<td>What causes Ebola?</td>
</tr>
<tr>
<td>Geography</td>
<td>13</td>
<td>Have there ever been cases in the UK?</td>
</tr>
<tr>
<td>Prevention</td>
<td>21</td>
<td>How to protect myself from Ebola?</td>
</tr>
<tr>
<td>Statistics</td>
<td>11</td>
<td>What is the mortality rate?</td>
</tr>
<tr>
<td>Symptoms</td>
<td>12</td>
<td>How would you know if you had it?</td>
</tr>
<tr>
<td>Transmission</td>
<td>39</td>
<td>How do people get Ebola?</td>
</tr>
<tr>
<td>Treatment</td>
<td>25</td>
<td>Is there a cure for Ebola?</td>
</tr>
<tr>
<td>Type</td>
<td>10</td>
<td>What is ebola rest on virus?</td>
</tr>
<tr>
<td>Excluded from analysis</td>
<td>5</td>
<td>Does it bother you that Rabies is more lethal than Ebola?</td>
</tr>
</tbody>
</table>

**Figure 2.** Frequency of question topics.

27.5% of the question totals (see **Figure 2**).

**4.2. Accuracy of Ebola Answers**

Overall, only 27.0% of the posted answers were rated as “accurate” (*i.e.* answering the question asked and containing no factual errors; see **Figure 3**). However, when accuracy was compared between answers to differing topics, substantial heterogeneity was observed, with between 11.8 - 45.5% of answers being rated as accurate (see **Table 2**). When Yahoo Answers’ “best answers” were examined, the overall accuracy was substantially higher, with 80.0% of “best answers” being rated as accurate compared to 16.0% of all other answers (see **Figure 4**).
4.3. Predictors of Answer Accuracy

Logistic regression modeling found that the overall model with all five predictors together served as a statistically significant predictor of answers’ accuracy ($\chi^2(5) = 25.08$, $p < 0.001$; Nagelkerke $R^2 = 0.37$). Examining the individual predictors revealed only a single statistically significant predictor of accurate answers (see Table 3). Specifically, answers that were voted “best answer” were approximately 21 times as likely to be rated accurate (OR = 21.32, 95% CI = 1.47 - 310.02, $p = 0.03$).

![Figure 3. Overall answer accuracy.](image)

<table>
<thead>
<tr>
<th>Question Type</th>
<th>n</th>
<th>Percent of Answers</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Accurate</td>
<td>Inaccurate</td>
</tr>
<tr>
<td>Biology</td>
<td>56</td>
<td>28.6</td>
</tr>
<tr>
<td>Geography</td>
<td>17</td>
<td>11.8</td>
</tr>
<tr>
<td>Causes</td>
<td>13</td>
<td>38.5</td>
</tr>
<tr>
<td>Prevention</td>
<td>21</td>
<td>28.6</td>
</tr>
<tr>
<td>Statistics</td>
<td>11</td>
<td>45.5</td>
</tr>
<tr>
<td>Symptoms</td>
<td>12</td>
<td>25.0</td>
</tr>
<tr>
<td>Transmission</td>
<td>39</td>
<td>23.1</td>
</tr>
<tr>
<td>Treatment</td>
<td>25</td>
<td>20.0</td>
</tr>
<tr>
<td>Type</td>
<td>10</td>
<td>40.0</td>
</tr>
</tbody>
</table>

Table 2. Percent of Answers in Each Accuracy Category by Type of Question Asked (n = 204).
Figure 4. Answer accuracy by “Best Answer” status.

Table 3. Summary of Logistic Regression Analysis for Variables Predicting Answer Accuracy (n = 81).

<table>
<thead>
<tr>
<th></th>
<th>B</th>
<th>SE</th>
<th>p</th>
<th>OR</th>
<th>95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Best Answer</td>
<td>3.06</td>
<td>1.37</td>
<td>0.03</td>
<td>21.32</td>
<td>1.47-310.02</td>
</tr>
<tr>
<td>Statistical Information</td>
<td>0.46</td>
<td>0.67</td>
<td>0.50</td>
<td>1.58</td>
<td>0.42-5.88</td>
</tr>
<tr>
<td>Source Disclosed</td>
<td>0.48</td>
<td>0.89</td>
<td>0.59</td>
<td>1.61</td>
<td>0.28-9.20</td>
</tr>
<tr>
<td>Link</td>
<td>0.82</td>
<td>1.02</td>
<td>0.42</td>
<td>2.27</td>
<td>0.31-16.60</td>
</tr>
<tr>
<td>Word Count</td>
<td>0.00</td>
<td>0.00</td>
<td>0.98</td>
<td>1.00</td>
<td>1.00-1.05</td>
</tr>
</tbody>
</table>

5. Discussion

Overall, the accuracy of Ebola information posted to Yahoo Answers was quite low, with less than half of all answers providing fully accurate information. More troubling, the questions that would be most relevant during an infectious disease outbreak, namely transmission, symptoms, and treatment, were each answered accurately less than a third of the time. In light of Internet users’ low electronic health literacy [6] [7], susceptibility to search biases [12] [14] [20], and tendency to base health behaviors off of online information [1] [13] [15]-[21] [23] [24], these data suggest that KESWs could serve as a source of misinformation and a driver of high risk behaviors during an infectious disease outbreak.

The finding that people who posted questions on the KESW later selected “best answers” that were 21 times more likely to be accurately answered helps to allay some of these concerns raised about visitors’ eHealth literacy. In aggregate,
it seems like KESWs users were able, to some degree, to discern accurate information from the various responses given. In fact, 80.0% of the answers voted “best answer” were accurate while only 2.9% of these answers were categorically inaccurate. That said there remain significant unknowns. For instance, while the users who posted the questions in this sample tended to select accurate “best answers”, it is unclear whether and how the demography of question posters might differ from users who only passively read through others’ questions and answers. In addition, it is unclear to what degree users rely on the best answers or instead read through multiple answers, possibly looking for the answer that most closely matches their preexisting beliefs or perceptions.

In addition, the observation that nearly a quarter of the responses represented attempts to troll the question poster speaks to the communities utilizing KESWs. Unlike health forums and support groups established to address a single health problem, KESWs appear to draw a more diverse population of Internet users, including Internet trolls. Anecdotally, several of the most egregious, inflammatory statements were attributable to a small number of repeat offenders whose inflammatory comments appeared under several questions. At the same time, only three of the answers provided came from users who indicated a relevant professional background.

Several limitations should be considered when examining these results. First, in the absence of further data, it is worth noting that the culture of KESW users may differ widely from Website to Website, limiting the generalizability of these findings. Further research is needed to explore not only how KESW users differ across different sites such as Yahoo Answers versus Reddit, but also how the culture of users differs across different health topics. For instance, the participation of vociferous groups like the anti-vaxxer community could radically change the distribution of accurate to inaccurate posted answers on topics like childhood vaccination recommendations. Likewise, it seems plausible that trolling may be more prevalent in posts related to topics being popularized by the media. Media coverage of infectious disease outbreaks may serve to draw trolls to posts related to those diseases.

Another limitation of this study is the treatment of accuracy as a dichotomous variable. The coding of websites’ accuracy has varied from study to study, with some evaluating the proportion of content that is accurate rather treating the content as either accurate or inaccurate. In this study, content with any misinformation at all was coded inaccurate, not only because of the potential harmful impact of any misinformation during an infectious disease outbreak, but also because misinformation surrounded by accurate information may be particularly insidious and difficult to detect. That said, examination of the ratio of accurate to inaccurate information within each KESW answer might be illuminating.

In addition, due the high ratio of answers to questions, although 204 answers were available to code, only 23 posts with 35 total questions were examined. This raises the possibility that other types of questions are being asked about Ebola on
KESWs, or that ratio of question topics being addressed may differ from those presented here. These data nonetheless take the first steps towards filling the knowledge gap on KESW answers’ accuracy, and future replication research will help to verify the types of questions being asked.

Ultimately, these data highlight the risks posed by seeking health information related to emerging infectious disease online through KESWs. Although those posting questions selected “best answers” that were often accurate, too little is known about the browsing habits of other KESW users. The presence of frequent misinformation among the posted responses and high volume of unhelpful information (unanswered, subjective, or trolling responses), suggest that these sites may pose special risks to users with low health literacy or medical misperceptions. In the context of Ebola, this misinformation could translate into challenges to outbreak containment, opposition to proper quarantine procedures, or social stigmatization of patients.

Further research is needed in order to explore the landscape of different KESWs and health topics, though these preliminary results raise concerns. If these patterns of inaccurate information hold true in other contexts, it may be necessary to provide users with tools to help them ascertain the veracity of user-generated claims, work directly with KESW providers to develop quality control mechanisms on their websites, and direct practitioners’ attention to these sites both to drive further research as well as to prepare practitioners to work with populations using these sites as a source of medical information.

References


[22] Holmes, M.M., Bishop, F.L. and Calman, L. (2017) “I Just Googled and Read Eve-
rhing”: Exploring Breast Cancer Survivors’ Use of the Internet to Find Information on Complementary Medicine. *Complementary Therapies in Medicine, 33*, 78-84.


https://doi.org/10.1097/INF.0000000000000755

[38] IBM Corp (2016) IBM SPSS Statistics for Mac, Version 24.0. IBM Corp., Armonk.