Uncertainty During New Disease Outbreaks in Wikipedia

Reham Al Tamime
University of Southampton
Southampton, UK
rat1g15@soton.ac.uk

Richard Giordano
University of Southampton
Southampton, UK
r.giordano@soton.ac.uk

Wendy Hall
University of Southampton
Southampton, UK
wh@ecs.soton.ac.uk

Abstract

Wikipedia relies on a community of editors who construct articles related to various topics such as new disease outbreaks. New disease outbreaks are often characterized by emergent and changing information which, in turn, require Wikipedia editors to spend time and effort to retrieve and understand information that is sometimes ambiguous, complex, and contradictory. This makes a good case study for identifying different types of uncertainty, as well as delineating strategies for managing uncertainty. Despite recent research that explores how the general population expresses uncertainty during crisis events on social media platforms, limited work has focused explicitly on how Wikipedia editors express uncertainty while editing articles when knowledge or background information is either incomplete or unknown. Thus, the goals of this study are to identify types of uncertainty expressed by Wikipedia editors during new disease outbreaks, and examine different strategies deployed by Wikipedia editors to manage uncertainty. Our findings may be used to build a theoretically and empirically-driven framework to better understand and manage uncertainty on Wikipedia.

Introduction

Uncertainty is defined as lack of predictability regarding a situation, individual, or behaviour (Berger and Calabrese 1975). Further research extended this definition by noting that "uncertainty exists when details of situations are ambiguous, complex, unpredictable or probabilistic; when information is unavailable or inconsistent; and when people feel insecure in their own state of knowledge or the state of knowledge in general" (Brashers 2001). Other researchers have argued that uncertainty is not something inherently negative, nor is it something always sought to be eliminated or resolved; rather, uncertainty is an object of ongoing appraisal and management (Brashers and Hogan 2013).

A disease outbreak is the occurrence of cases of disease in excess of what would normally be expected in a defined community, geographical area or season. An outbreak may occur in a restricted geographical area, or may extend over several countries. It may last for a few days or weeks, or for several years (World Health Organization 2018). Optimal decision making during a disease outbreak or the emergence of a new pathogen is often hampered by considerable uncertainty due to the lack of knowledge about the disease etiology, transmission, contagion, morbidity, or mortality. Therefore, decisions are usually informed through retrospective analyses of prior crises, trials, and interventions (Shea et al. 2014). The high level of uncertainty about different aspects of disease outbreaks, and the manner by which they are spread by the media, sometimes stigmatize those infected and lead to deep concern and fear among the well. (Osorio-de Castro et al. 2017).

Wikipedia has become the most prominent source of online health information for the general public compared to other online health information providers, such as MedlinePlus or NHS Direct (Laurent and Vickers 2009). Although research has been conducted to examine editing dynamics and page views on Wikipedia during crises (Keegan, Gergle, and Contractor 2011; Priedhorsky et al. 2017), there is limited research that focuses on how contributors to Wikipedia experience and manage uncertainty during new disease outbreaks. Therefore, our research questions are, *What are the different types of uncertainty expressed by Wikipedia editors during a crisis such as a new disease outbreak? How do Wikipedia editors manage different forms of uncertainty during a disease outbreak?* This qualitative study on Wikipedia talk pages has been conducted to address these research questions. This work offers some understanding of the types of uncertainty that emerges when collaborating editors construct encyclopedic knowledge during a real-time crisis, as well as the strategies they use to manage uncertainty.

Related Work

Health and Wellbeing on the Web

The use of social media and search engines among public health organizations and patients has proliferated over the last years (Antheunis, Tates, and Nieboer 2013). Health organizations mainly use social networking applications to build relationships and to provide medical information to the public (Antheunis, Tates, and Nieboer 2013; Ndumbe-Eyoh and Mazzucco 2016), while patients use social media sites to fill a need for health information and receive advice, social and emotional support (Zhao and Zhang 2017). Research has indicated that people modify their information seeking and sharing practices on search engines or social media depend-

Copyright © 2019, Association for the Advancement of Artificial Intelligence (www.aaai.org). All rights reserved.

ing on whether their condition is serious, highly stigmatic, or insignificant (De Choudhury, Morris, and White 2014).

The widespread use of online applications for health and wellbeing has led researchers to explore the characteristics of different online channels and communities. Chomutare et al. (2013) studied the dynamics of an online community for diabetes patients, and found that the most centric members (those with more trust within the communities) tend to have more years of experience with the disease. Other studies found that members of the TuDiabetes online community often construct shared meaning through deep discussions, back and forth negotiation of perspectives, and resolution of conflicts (Mamykina, Nakikj, and Elhadad 2015). Different studies looked at the perceived satisfaction of members of an online breast cancer community (Vlahovic et al. 2014). They revealed that women's satisfaction—the effect experienced when expectation-type standards are fulfilled (Hecht 1978)—depends on whether or not they retrieve what they seek. Specifically, women were less satisfied if they sought information and instead received emotional support in return. Yoo, Kim, and Lee (2018) disclosed that social media-related perceptions (i.e., informational subjective norms on social media, channel beliefs, social media efficacy) have an indirect impact on behavioral intention through communicative behaviors.

Several studies stressed the benefits of using social media and search engines for health and wellbeing. For example, sharing health data among members of an online community was found to be correlated with improved personal management of the disease (Santoro et al. 2015). Also, Facebook-based intervention approaches can increase physical activity among young adult cancer survivors (Santoro et al. 2015). Similar studies showed that patients with cardiovascular diseases felt that using social media in health education would improve the efficiency and quality of health education programs, and lead to better outcomes (Eshah 2018). Other studies explored the potential of mining social media data for public health surveillance and monitoring such as detection of disease outbreaks and awareness, pharmacovigilance, and various issues related to behavioral medicine, including weight loss, and e-cigarette use (Paul et al. 2016; Eggleston and Weitzman 2014).

Crises and the Web

Crises—periods of perceived intense difficulty or danger—are occasions of high uncertainty and threat that call for swift attention and decision making. The terms *emergency*, *crisis*, *disaster* and *catastrophe* are the key words used to capture the scale and significance of such episodes (Eriksson and McConnell 2011). During crises, the use of social media such as Facebook and Twitter increases (Austin, Fisher Liu, and Jin 2012). The public uses social media during crises to seek neighborhood-specific interactive fora, reconnect with others through social networks, gather information on property damage, transfer, translate, and transform knowledge, monitor information, spread awareness, express concerns and offer help to others (Finch et al. 2016; Austin, Fisher Liu, and Jin 2012).

Researchers studied the characteristics of information generated during crises on different social media and online forums. Vieweg et al. (2010) analyzed Twitter data during the Spring 2009 Red River Floods and Oklahoma Grass Fires and identified different features related to situational updates including warning, preparatory activity, fire-line or hazard location, flood level, weather, wind, visibility and road conditions. Additional analysis of Twitter data during the Spring 2009 Red River Floods revealed that keyword-containing re-tweets from geographically local people are more likely to pertain to the event (Palen et al. 2010). Moreover, Qu, Wu, and Wang (2009) found that local citizens used a popular online forum to organize information, express their emotions and to provide support to one another during the 2008 Sichuan Earthquake.

Other researchers use social media data to examine how individuals and groups experience uncertainty and risk during crises. Starbird et al. (2014) studied the propagation of misinformation online in the aftermath of the Boston Marathon Bombings. They found that corrections to misinformation emerge, but are muted compared with the propagation of the misinformation. Also, Starbird et al. (2016) identified and classified different types of uncertainty expressed by individuals on Twitter during the Boston Marathon Bombings and the Sydney Siege crises. They showed that milling behaviours, such as behaviours related to making sense of uncertain spaces including interpreting, speculating, theorizing, debating, or challenging, were prominent in Twitter posts during the events. Other work applied qualitative content analysis to investigate how individuals used social media in response to the Zika virus outbreak (Gui et al. 2017). While focusing on travel-related decisions, they noticed that the general population experienced extreme uncertainty during the Zika virus crisis. This uncertainty arose as a consequence of several gaps in authoritative knowledge as well as lack of scientific knowledge about the virus. Similar studies examine the public's communication and perception of risks on social media during the Zika crisis (Gui et al. 2018). The researchers recognized two types of perceived risks: ones directly caused by the Zika virus; and ones introduced by authorities' risk management measures.

Crises Response and the Web

In crisis situations, assessing and responding to the humanitarian needs of the affected population is a critical task. The response efforts involve coordinating with agencies to provide aid and help as well as find and evacuate victims (Dugdale, Van de Walle, and Koeppinghoff 2012). Social media tools have changed the landscape of crisis management considerably over recent years and allowed more possibilities to manage and respond to crises. Citizens, governments, non-governmental organizations (NGOs) and others often adopt a *connect and collaborate* approach through the use of social media during crises (Akhgar et al. 2013).

Social media applications may afford early detection of potential disease outbreaks and improved tracking of disease and injury trends (Finch et al. 2016). Social media applications also allow public health workers and emergency responders to act quickly and efficiently during crises (Finch

et al. 2016). Social media enables health-care agencies and governments to share data with their citizens, and ascertain how public sentiments, rumors, and information (false and otherwise) are spread. Consequently, agencies can target public health messages and deploy effective interventions (Kass-Hout and Alhinnawi 2013). Accordingly, different applications have been developed to integrate information coming from social media during crises. Social media applications, in general, have been found to contribute to the improvement of communication prior to and during a crisis event (Maresh-Fuehrer and Smith 2016). For example, The Ushahidi Platform is an open source web application for information collection, visualization and interactive mapping. It allows people to collect and share their own stories using various media such as SMS, Web Forms, Email or Twitter (Dugdale, Van de Walle, and Koeppinghoff 2012). Other applications such as SituMap and PhotoSorter were designed to include features that expand the communication options available during a crisis. These features enabled a timely and effective response during a crisis, as well as the capability to track weather, traffic, and other relevant information (Maresh-Fuehrer and Smith 2016).

Other work explored the application of crowdsourcing to effectively respond to and manage crises. Crowdsourcing in a crisis context is the use of ICTs to mobilize volunteers across the globe, report crisis situations from the ground, translate reported messages, carry out crisis mapping, and self-organize the coordination of relief resources (Park and Johnston 2015). For example, Starbird (2011) used the Twitter platform for crowdsourcing information during disasters and mass emergency events by asking Twitter users to incorporate special hashtags into their crisis-related tweets to make these tweets machine-readable. Other researchers developed a prototype of City-Share, which relies on crowd-based crisis management and situated nature of voluntary activities (Ludwig et al. 2017). Other work explored the mapping practices and the interactions of the Open Street Map volunteers during crises (Kogan et al. 2016) and identified seven distinct mapping practices that can be classified according to dimensions of time, space, and interpersonal interaction.

Research studies assessed the use of different social media applications for crises management. Stewart and Gail Wilson (2015) outlined both the bright- and dark-side uses of social media by emergency respondents during the Hurricane Sandy crisis. The bright side includes listening to an audience and engaging conversation. On the other hand, the dark side includes the dissemination of rumors, hoaxes and misinformation. Other researchers studied how the Center for Disease Control and Prevention (CDC) used Twitter during the Ebola outbreak (Dalrymple, Young, and Tully 2016). They found that although the CDC's tweets reflected extant protocols and contained facts about disease transmission, the CDC had a contradictory role of conveying facts and managing public fear in a chaotic and uncertain situation. These conditions, they argue, can make any authoritative organization appear to the public to be both unprepared and unclear.

Health, Crises, and Crises Management on Wikipedia

Wikipedia affords both health and non-health professionals the opportunity to contribute to its medical and health content. The English-language Wikipedia is a prominent source of online health information compared to other providers such as MedlinePlus and NHS Direct (Laurent and Vickers 2009). Editors are motivated to contribute to health-related content for various reasons including education, help, improving their own clinical practice (when relevant) and personal intellectual enjoyment (Farič and Potts 2014).

The use of Wikipedia during crises has been recognized in previous research. Keegan, Gergle, and Contractor (2011) selected the T¯ōhoku catastrophe article to examine how the Wikipedia community responds to unexpected events and emergencies. The study identified several dynamic features of collaboration during unexpected events. These features include intense editing activities concerning articles related to these events. Tausczik et al. (2012) used the number of daily visits to H1N1-relevant Wikipedia articles to assess anxiety and information-seeking behaviours in response to the 2009 H1N1 outbreak. The number of visits increased after the announcement of the outbreak of H1N1 and then declined rapidly. Al Tamime, Giordano, and Hall (2018) delineated and classified different levels of editorial *burstiness* in Wikipedia during new disease outbreaks. Other researchers explored the potential to use the information on Wikipedia for crises management. For example, the number of daily visits to Wikipedia articles has been used by Generous et al. (2014) to build a model that can monitor and forecast the incidence of infectious disease. Despite the study's limitations, the model was successful in monitoring and forecasting the progress of infectious disease over the course of up to 28 days. Nevertheless, recent research by Priedhorsky et al. (2017) that evaluated the use of Wikipedia traffic data and category links to monitor and forecast diseases found very little forecasting value. Instead, the researchers argued that the field of internet-based disease measurement does not only need experiments and simulation, but also mathematical theory to describe the flow of disease-related information from human observations through internet systems and algorithms to actionable knowledge.

While studies of health and wellbeing on the web mainly concern on the use of social media by public health organizations and patients, our study is instead focused on the use of Wikipedia by editors, particularly for building health-related information. Different from research that explores different forms of uncertainty expressed by the public during crises, this research scrutinizes different forms of uncertainty expressed by Wikipedia editors during crises. Compared to other work that scouts the use of social media for crises response and management, this work examines different strategies employed by Wikipedia editors to manage information uncertainty during crises. Our work builds on these research strands by using qualitative analysis to elucidate how Wikipedia editors experience and cope with uncertainties during new disease outbreaks.

Methods

We applied qualitative inductive content analysis on Wikipedia talk pages. The aim of content analysis is to attain a condensed and broad description of the phenomenon, and the outcome of the analysis is usually concepts or categories describing the phenomenon (Elo and Kyngäs 2008). We selected inductive content analysis as a method because former qualitative research and knowledge about editing practices during uncertainty are limited. We examined Wikipedia talk pages to gain insights about the process of editing Wikipedia articles during new disease outbreaks.

Data

We studied the talk pages of five new disease outbreaks that occurred in the past ten years. These new diseases are the Zika virus, Ebola, Swine Influenza, Avian Influenza (H7N9) and Middle East Respiratory Syndrome (MERS). The talk page archives of these articles were collected, covering all the content from 2001 to 2018. This produced a total of 261 talk pages.

Data Analysis

The analysis of the talk pages was done in stages. In the first stage, the authors conducted an initial round of coding on all the sampled talk pages to find several categories and subcategories about how editors experienced uncertainty, and how they managed it. The initial coding generated two main recurring themes: 1) types of uncertainty; and 2) strategies to handle uncertainty. In the second stage, the authors read through the data again and generated sub-themes under the two main themes. Through several rounds of reading, coding and comparing emerging themes, we created several subthemes to describe the forms of uncertainty expressed by editors during new disease outbreaks, as well as different strategies employed to manage uncertainty.

Results

In this section, we report two themes that emerged from our content analysis of Wikipedia talk pages during new disease outbreaks. We first discuss sources of uncertainty that the editors perceived while working on articles during new disease outbreaks. We then report the strategies used to cope with different forms of uncertainty.

Types of Uncertainty

The analysis revealed that there are three types of uncertainty perceived by editors during new disease outbreaks: scientific uncertainty (the most prominent), conflicting references uncertainty, and reference uncertainty. These are discussed in turn.

Scientific uncertainty Scientific uncertainty is expressed in Wikipedia talk pages by editors as a result of doubt and ambiguity about scientific knowledge related to diseases. These doubts are related to the etiology, transmission or treatment of diseases and emerge as a consequence of either poor or no scientific evidence or publications about the diseases during their outbreaks. For example, an editor expressed scientific uncertainty during the Zika virus outbreak by commenting:

> *Zika virus has been grabbing headlines because of its links to an alarming birth defect called microcephaly. The data to provide evidence linking the relatively mild mosquito-borne disease and babies born with small heads and potential brain damage, however, are not yet conclusive.*

Other editor replied to a question about the immunity of Ebola survivors by expressing uncertainty because of a lack of evidence:

> *I suspect that the experience and information is still, sadly lacking. My concerns are: Lack of evidence of effective protection from this virus from previous strains, lack of evidence that immunity to re-exposure to this strain is harmless due to immunity for the first quarter second thoughts. The list continues for far longer lengths within the second enough to become absurd, due to the lack of knowledge on the first two consideration...*

We can see from the above examples that scientific uncertainty is accompanied by phrases to express doubt such as "not yet conclusive" and "lack of evidence". As shown in Table 1, scientific uncertainty is the most common type of uncertainty expressed on Wikipedia talk pages during new disease outbreaks.

Conflicting references uncertainty Conflicting references uncertainty emerge on Wikipedia talk pages as a result of different references showing contradictory information about diseases. Editors use the Wikipedia talk pages to clarify or defend their points of view as conflicting references uncertainty might lead to controversy between editors. An editor stated during the H7N9 outbreak:

> *Is it necessary to have a tabulated breakdown of total number of cases in each provincial-level areas in China? I may do it but an inconsistency problem may exist for the data provided by the websites of the health departments of each affected province do not match with that of Xinhua.*

This shows that Wikipedia editors try to compare different references to track the number of cases during outbreaks. Nevertheless, it might be difficult to find consistency among different reporting agencies–some of the inconsistency can be explained by the timing and release of information or research. For example, an editor replied to a question about the origin of the virus during the swine influenza outbreak:

> *That was the finding of the very first analyses of the sequences, as reported in reliable sources. But now there is a newer analysis that gets different results...*

This suggests both the changing nature of certain scientific knowledge in general and certain scientific knowledge during a crisis in particular.

References uncertainty References uncertainty refers to doubts and ambiguity about information or claims within the

	Scientific uncertainty	Conflicting references	Reference uncertainty
Ebola	29%	24%	9%
H7N9	7%	2%	0
MERS	0	0	0
Swine influenza	9%	2%	0
Zika virus	18%	0	0
New Disease Outbreaks	62%	29%	9%

Table 1: Uncertainty Type as a Percentage of Overall Uncertainty

references themselves. Editors are responsible to find references and link them to Wikipedia articles, and editors sometimes seek clarification on the content of some references or consult other editors to see if certain references are valid. For example, one of the editors questioned a reference during the Ebola outbreak:

> *I find this paper fascinating. . . They seem to believe the 1976 outbreak originated in Sudan. But I thought the Sudan and Zaire 1976 outbreaks ended up being two separate species so that means there were two separate index patients two months apart and about 1000 km apart? But that is also very unlikely which is why I see that these researchers thought it came from Sudan.*

Another editor questioned a claim from the CDC and WHO:

> *Both WHO and the CDC say that the incubation period for Ebola is 2 to 21 days. How did they come up with 21 days? I can't seem to find it in this or related Wikipedia articles.*

In the last example, some uncertainties were about references that belong to authoritative sources such as the World Health Organisation (WHO) and Centers for Disease Control and Prevention (CDC). Also, the Ebola article was the only article that included forms of references uncertainty.

Strategies to Manage Uncertainty

Posting sources of uncertainty on Wikipedia talk pages to discuss them with other editors is the first step to handle uncertainty. Here we explore these posts to outline the different strategies that have been used by Wikipedia editors to cope with uncertainty during new disease outbreaks.

Rely on authoritative sources Information coming from different sources might confuse editors and increase their level of uncertainty. Therefore, editors try to rely on the CDC and the WHO as authoritative sources to clear doubts during disease outbreaks. For instance, some editors found conflicting information about an anti-viral drug used for swine influenza treatment, so one of them wrote:

> *Looks like what is on the CDC website agrees with this article but not that press story. I'll stick with what the CDC say on the matter.*

However, statements coming from the CDC and the WHO might be perceived by editors as contradicting each other. For example, an editor had a question about conflicting information about the transmission of Ebola coming from the CDC and the WHO. Another editor replied to his question saying:

> *Have you written to the CDC to ask them to clarify?*

The above post also suggests that relying on published authoritative sources might not be a straightforward task during disease outbreaks as editors might find conflicting information from such published sources. Instead, they may need to check with the publishing organizations.

Report the uncertainty in the article During new disease outbreaks, editors might not be able to find definitive answers to all of their questions. In such cases, a strategy they use is to inform readers of Wikipedia articles about such uncertainties explicitly. For example, one of the editors commented during the H7N9 outbreak:

> *How about the following for a sentence in the intro paragraph: Researchers have commented on the unusual prevalence of older males among H7N9-infected patients. While several environmental, behavioral, and biological explanations for this pattern have been proposed, as yet, the reason for this is not known.*

Other editors tried to find phrases to tone down the language and to make the information sounds less definitive. During the Ebola outbreak, an editor posted:

> *Until they find the reservoir species with sufficient quantities of whole virus and can prove the animal-human infection cycle, I think any statement needs to contain a "most likely" hedging statement or similar.*

Ignore the uncertainty In other cases, mentioning the uncertainty may result in the article's becoming unduly complex, so the editorial decision is not to mention the uncertainty at all and, in effect, editing it away until more certain knowledge emerges. For example, when one editor had a question about a specific type of treatment of Ebola, he received this answer on the talk page:

> *If you cannot find any literature or sources supporting this then by all means, remove it from the main page and place it on the talk page here so future editors can try to reference it.*

In addition, an editor found various sources reporting different dates of suspected first case during the Ebola outbreak:

> *I think we should just say "December 2013" and not try to put a date on it, because the date within December doesn't appear well-established*

These examples prove that some editors try to provide general information instead of focusing on uncertain and maybe confusing details.

	Rely on authoritative sources	Report the uncertainty	Ignore the uncertainty	Consult experts for advice	Set up mailing list
Ebola	19%	4%	11%	11%	0
H7N9	0	7%	11%	0	0
MERS	0	0	0	0	0
Swine influenza	7%	0	0	0	0
Zika virus	7%	15%	0	4%	4%
New Disease Outbreaks	33%	26%	22%	15%	4%

Table 2: Strategies to Manage Uncertainty as a Percentage of Overall Number of Strategies

Consult experts for advice Different editors contact domain experts to clarify some information during disease outbreaks. These experts could be doctors or health professionals who are also involved in editing health-related articles. For example, to address the uncertainty regarding the immunity of Ebola survivors, an editor commented:

> *Does anyone know an immunologist so we can get this right?*

Another editor replied:

> *I put in a request for assistance at wikipedia project medical. I suspect that the experience and information is still, sadly lacking.*

This post highlights that finding experts to clarify information is not an easy task during disease outbreaks as experts might also be uncertain about different aspects of the diseases.

Set up a mailing list about ongoing research Some editors take the initiatives to identify sources of uncertainty and track ongoing research that addresses this uncertainty. Specifically, an editor indicated during the Zika virus outbreak:

> I am setting up a mailing list (ttps://groups.google.com/forum/#!forum/zika-researc) for ongoing research around Zika virus, Zika fever and Zika virus outbreak (2015 – present). So if any questions come up here that need expert input, please ping me or post there directly.

Discussion

This work suggests that different types of uncertainties emerge as editors deal with ambiguous, incomplete, or contradictory information when constructing Wikipedia articles on disease outbreaks. We also show that there are a number of strategies that they use to manage these uncertainties, such as relying on authoritative sources and consulting experts. If the uncertainty cannot be resolved, they will either report the uncertainty in the article, or ignore the uncertainty if reporting it would add unneeded complexity to the article. In a rare case, an editor will set up a mailing list to gather the results of ongoing research that can be used to resolve the uncertainty.

Multiple Forms of Uncertainty

Various forms of uncertainty are conveyed by Wikipedia editors during a disease outbreak. These forms exist at the scientific and editorial levels. At the scientific-level, uncertainties are related to doubt and ambiguity about scientific content and knowledge related to diseases. In contrast, editorial-level uncertainties are related to doubt and ambiguity about references used to build Wikipedia articles. Moreover, Wikipedia editors needed clarification on statements originating from international public health agencies. Different forms of uncertainty are often multi-layered and interconnected, leading to increased levels of uncertainty among editors. For instance, scientific uncertainty might also lead to references or conflicting references uncertainty because incomplete research and untested knowledge about diseases might make it difficult to clarify information during disease outbreaks.

Brashers (2001) suggested that uncertainty is likely to vary across contexts; various dimensions of uncertainty suggest that individuals develop responses sensitive to multiple goals and tasks of information seeking. Previous research studied how uncertainty is expressed on social media and online fora in different contexts, such as uncertainty associated with rumours (Starbird et al. 2016) and uncertainty associated with travel-related decisions (Gui et al. 2017). Starbird et al. (2016) identified several milling behaviours—people gathering in times of crisis to discuss and attempt to understand the cause of an event (Turner 1964)—in rumoring Tweets during crisis events. These milling behaviors include challenging and doubting a rumor, asking both leading and open questions either to seek information, questioning the source or the credibility of the rumor, communicating uncertainty and contradictory information about rumours, and expressing incredulity, hope or fear. In addition, Gui et al. (2017) found that people express uncertainty on Reddit, TripAdvisor and BabyCenter as they perceive authoritative information to be incomplete, inaccurate, and insufficient to make travel decisions during the Zika outbreak. Posts that express uncertainty include overgeneralized and inaccurate reporting of information, questioning or expressing distrust in authoritative sources, reporting inconsistent information from healthcare providers, and questioning whether the media are reporting information in an objective manner. A study has delineated evidence of expressed doubts on authentic reports and rumors spreading in the Tianya online forum after 2008 Sichuan Earthquake (Qu, Wu, and Wang 2009).

While these studies looked at different forms of uncertainty expressed by the public during crises, our work instead examines different forms of uncertainty expressed

Article	Number of editors in Wikipedia edit history	Number of editors in Wikipedia talk page	Number of editors in talk page as a percentage of number of editors in edit history
Ebola	3566	415	12%
H7N9	219	31	14%
MERS	351	28	8%
Swine influenza	1589	433	27%
Zika virus	559	76	14%

Table 3: Proportion of Editors Participating in on Sample Talk Pages

by Wikipedia editors when building articles during public health crises. Our findings agree with prior studies that Wikipedia editors also find information to be incomplete, inaccurate and contradictory during outbreaks. Wikipedia editors also question information coming different sources, including authoritative sources. Our findings, however, also delineate the strategies that editors use to manage uncertainties.

Strategies to Manage Uncertainty

Wikipedia editors depend on several strategies to cope with uncertainty during a disease outbreak. These strategies rely primarily on consulting authoritative sources, reporting the uncertainty to the public, ignoring the uncertainty in the interests of maintaining simplicity, and, to a far lesser extent, setting up a mailing list to gather information and science as they emerge over time. These strategies are sometimes discrete or used in combination with each other.

Brashers (2001) identified several strategies to manage uncertainty in illness such as seeking information and adapting to chronic uncertainty. Information seeking behavior is used to decrease uncertainty by attempting to fill gaps in information or knowledge, or to confirm/disconfirm the current belief state (Brashers 2001). We found that editors attempt to reduce uncertainty by seeking information from authoritative published sources, by asking for experts' advice, as well as by following ongoing research on disease outbreaks. Accepting uncertainty is an adaptive mechanism used by editors to deal with ongoing uncertainty. Adapting to chronic uncertainty can involve several behaviours such as ignoring the uncertainty-producing event (in the interests of maintaining simplicity) or altering planning and decision-making strategies (Brashers 2001). Therefore, we found that editors can sometimes adapt to uncertainty by editorially ignoring the uncertainty in the interests of an article's clarity.

Previous research focused on information seeking as a strategy to manage uncertainty during crises and disease outbreaks. For example, Vieweg et al. (2010) reported that different audiences on the individual as well as on the community level may seek information and situational updates from Twitter data during uncertain events. Similarly, Qu, Wu, and Wang (2009) concluded that that Tianya online forum members tried to clarify rumors and resolve doubts by seeking further information, cross-validating with other people's personal experience, and comparing information with official announcements and news. Gui et al. (2017) showed that people turned to online forums to seek information that they could not obtain from authoritative sources, including local information and alternative authoritative information. Similar research also examined strategies employed by the CDC to manage uncertainty in Twitter during Ebola outbreak (Dalrymple, Young, and Tully 2016). Although engaging with the public was identified as a strategy to manage uncertainty, the public were not seeking engagement in the sense of dialogue or conversation. Instead, the public were probing for concrete information to assess personal risk. Our study agrees with previous findings that Wikipedia editors tend to also seek information to handle uncertainty by asking experts for advice, relying on information coming from authoritative sources, and keeping an eye on ongoing research. However, our study extends previous research that focused only on information seeking a strategy to handle uncertainty by revealing strategies that include ignoring the uncertainty and reporting the uncertainty in articles.

Limitations

While previous research focused on one or two crises to study, we have instead examined five different new disease outbreaks to understand how editors experience and manage uncertainty. However, our findings do not represent the views of *all* Wikipedia editors because not all editors participate in talk pages. As shown in Table 3, only a subset of Wikipedia editors posts in Talk pages and therefore, only this subset has been considered for this study. In addition, we have selected only a subset of new disease outbreaks. Finally, we have included only those articles that were written and edited in English.

Implications and Future Work

The primary goals of this work are to identify different types of uncertainty, and to delineate how uncertainty is managed by editors in the construction of a health-related Wikipedia article. We identify distinct types of uncertainty, along with strategies to cope with uncertainty. Our work has so far focused on new disease outbreaks where information may be incomplete or contradictory. Is this a feature only of new disease outbreaks, or is uncertainty or controversy a feature of other types of health-related articles, and to what extent can trusted sources be trusted? (Moreira 2007). We therefore intend to expand our analyses to chronic diseases and vaccinations. Further analysis could help in building a theoretically and empirically-driven framework to better understand and manage uncertainty on Wikipedia. We are

also looking to investigate the relationship between uncertainty and controversy—manifest disagreement between editors—in Wikipedia. We are also studying the potential of automatically analyzing written text to detect levels of uncertainty.

Acknowledgments

This work was supported by Qatar Foundation and the EPSRC Centre for Doctoral Training in Web Science Innovation under grant number EP/L016117/1.

References

Akhgar, B.; Fortune, D.; Hayes, R. E.; Guerra, B.; and Manso, M. 2013. Social media in crisis events: open networks and collaboration supporting disaster response and recovery. In *2013 IEEE International Conference on Technologies for Homeland Security (HST)*, 760–765.

Al Tamime, R.; Giordano, R.; and Hall, W. 2018. Observing burstiness in Wikipedia articles during new disease outbreaks. In *Proceedings of the 10th ACM Conference on Web Science - WebSci '18*, 117–126. Amsterdam, Netherlands: ACM Press.

Antheunis, M. L.; Tates, K.; and Nieboer, T. E. 2013. Patients' and health professionals' use of social media in health care: motives, barriers and expectations. *Patient Education and Counseling* 92(3):426–431.

Austin, L.; Fisher Liu, B.; and Jin, Y. 2012. How audiences seek out crisis information: exploring the social-mediated crisis communication model. *Journal of Applied Communication Research* 40(2):188–207.

Berger, C. R., and Calabrese, R. J. 1975. Some explorations in initial interaction and beyond: toward a developmental theory of interpersonal communication. *Human Communication Research* 1(2):99–112.

Brashers, D. E., and Hogan, T. P. 2013. The appraisal and management of uncertainty: Implications for information-retrieval systems. *Information Processing & Management* 49(6):1241–1249.

Brashers, D. E. 2001. Communication and uncertainty management. *Journal of Communication* 51(3):477–497.

Chomutare, T.; Årsand, E.; Fernandez-Luque, L.; Lauritzen, J.; and Hartvigsen, G. 2013. Inferring community structure in healthcare forums: an empirical study. *Methods of Information in Medicine* 52(2):160–167.

Dalrymple, K. E.; Young, R.; and Tully, M. 2016. "Facts, not fear": negotiating uncertainty on social media during the 2014 Ebola crisis. *Science Communication* 38(4):442–467.

De Choudhury, M.; Morris, M. R.; and White, R. W. 2014. Seeking and sharing health information online: comparing search engines and social media. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '14, 1365–1376. New York, NY, USA: ACM.

Dugdale, J.; Van de Walle, B.; and Koeppinghoff, C. 2012. Social media and SMS in the Haiti earthquake. In *Proceedings of the 21st International Conference on World Wide Web*, WWW '12 Companion, 713–714. New York, NY, USA: ACM.

Eggleston, E. M., and Weitzman, E. R. 2014. Innovative uses of electronic health records and social media for public health surveillance. *Current Diabetes Reports* 14(3).

Elo, S., and Kyngäs, H. 2008. The qualitative content analysis process. *Journal of Advanced Nursing* 62(1):107–115.

Eriksson, K., and McConnell, A. 2011. Contingency planning for crisis management: recipe for success or political fantasy? *Policy and Society* 30(2):89–99.

Eshah, N. F. 2018. Investigating cardiovascular patients' preferences and expectations regarding the use of social media in health education. *Contemporary Nurse* 54(1):52–63.

Farič, N., and Potts, H. W. 2014. Motivations for contributing to health-related articles on Wikipedia: an interview study. *Journal of Medical Internet Research* 16(12):e260.

Finch, K. C.; Snook, K. R.; Duke, C. H.; Fu, K.-W.; Tse, Z. T. H.; Adhikari, A.; and Fung, I. C.-H. 2016. Public health implications of social media use during natural disasters, environmental disasters, and other environmental concerns. *Natural Hazards* 83(1):729–760.

Generous, N.; Fairchild, G.; Deshpande, A.; Del Valle, S. Y.; and Priedhorsky, R. 2014. Global disease monitoring and forecasting with Wikipedia. *PLoS Computational Biology* 10(11):e1003892.

Gui, X.; Kou, Y.; Pine, K. H.; and Chen, Y. 2017. Managing uncertainty: using social media for risk assessment during a public health crisis. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems - CHI '17*, 4520–4533. Denver, Colorado, USA: ACM Press.

Gui, X.; Kou, Y.; Pine, K.; Ladaw, E.; Kim, H.; Suzuki-Gill, E.; and Chen, Y. 2018. Multidimensional risk communication: public discourse on risks during an emerging epidemic. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, CHI '18, 214:1–214:14. New York, NY, USA: ACM.

Hecht, M. L. 1978. Measures of communication satisfaction. *Human Communication Research* 4(4):350–368.

Kass-Hout, T. A., and Alhinnawi, H. 2013. Social media in public health. *British Medical Bulletin* 108(1):5–24.

Keegan, B.; Gergle, D.; and Contractor, N. 2011. Hot off the wiki: dynamics, practices, and structures in Wikipedia's coverage of the Tōhoku catastrophes. In *Proceedings of the 7th International Symposium on Wikis and Open Collaboration - WikiSym '11*, 105. Mountain View, California: ACM Press.

Kogan, M.; Anderson, J.; Palen, L.; Anderson, K. M.; and Soden, R. 2016. Finding the way to OSM mapping practices: bounding large crisis datasets for qualitative investigation. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, CHI '16, 2783–2795. New York, NY, USA: ACM.

Laurent, M. R., and Vickers, T. J. 2009. Seeking health information online: does Wikipedia matter? *Journal of the American Medical Informatics Association* 16(4):471–479.

Ludwig, T.; Kotthaus, C.; Reuter, C.; Dongen, S. v.; and Pipek, V. 2017. Situated crowdsourcing during disasters: managing the tasks of spontaneous volunteers through public displays. *International Journal of Human-Computer Studies* 102:103–121.

Mamykina, L.; Nakikj, D.; and Elhadad, N. 2015. Collective sensemaking in online health forums. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems - CHI '15*, 3217–3226. Seoul, Republic of Korea: ACM Press.

Maresh-Fuehrer, M. M., and Smith, R. 2016. Social media mapping innovations for crisis prevention, response, and evaluation. *Computers in Human Behavior* 54:620–629.

Moreira, T. 2007. Entangled evidence: knowledge making in systematic reviews in healthcare. *Sociology of Health & Illness* 29(2):180–197.

Ndumbe-Eyoh, S., and Mazzucco, A. 2016. Social media, knowledge translation, and action on the social determinants of health

and health equity: a survey of public health practices. *Journal of Public Health Policy* 37(S2):249–259.

Osorio-de Castro, C.; Silva Miranda, E.; Machado de Freitas, C.; Rochel de Camargo, K.; and Cranmer, H. H. 2017. The Zika virus outbreak in Brazil: knowledge gaps and challenges for risk reduction. *American Journal of Public Health* 107(6):960–965.

Palen, L.; Starbird, K.; Vieweg, S.; and Hughes, A. 2010. Twitter-based information distribution during the 2009 Red River Valley flood threat. *Bulletin of the American Society for Information Science and Technology* 36(5):13–17.

Park, C. H., and Johnston, E. 2015. Crowdsourced, voluntary collective action in disasters. In *Proceedings of the 16th Annual International Conference on Digital Government Research*, dg.o '15, 329–330. New York, NY, USA: ACM.

Paul, M. J.; Sarker, A.; Brownstein, J. S.; Nikfarjam, A.; Scotch, M.; Smith, K. L.; and Gonzalez, G. 2016. Social media mining for public health monitoring and surveillance. In *Biocomputing 2016*, 468–479. Kohala Coast, Hawaii, USA: WORLD SCIENTIFIC.

Priedhorsky, R.; Osthus, D.; Daughton, A. R.; Moran, K. R.; Generous, N.; Fairchild, G.; Deshpande, A.; and Del Valle, S. Y. 2017. Measuring Global Disease with Wikipedia: Success, Failure, and a Research Agenda. *CSCW : proceedings of the Conference on Computer-Supported Cooperative Work. Conference on Computer-Supported Cooperative Work* 2017:1812–1834.

Qu, Y.; Wu, P. F.; and Wang, X. 2009. Online community response to major disaster: a study of Tianya forum in the 2008 Sichuan earthquake. In *2009 42nd Hawaii International Conference on System Sciences*, 1–11.

Santoro, E.; Castelnuovo, G.; Zoppis, I.; Mauri, G.; and Sicurello, F. 2015. Social media and mobile applications in chronic disease prevention and management. *Frontiers in Psychology* 6.

Shea, K.; Tildesley, M. J.; Runge, M. C.; Fonnesbeck, C. J.; and Ferrari, M. J. 2014. Adaptive management and the value of information: learning via intervention in epidemiology. *PLoS Biology* 12(10):e1001970.

Starbird, K.; Maddock, J.; Orand, M.; Achterman, P.; and Mason, R. 2014. Rumors, false flags, and digital vigilantes: misinformation on Twitter after the 2013 Boston Narathon Bombing. In *iConference 2014 Proceedings*. iSchools.

Starbird, K.; Spiro, E.; Edwards, I.; Zhou, K.; Maddock, J.; and Narasimhan, S. 2016. Could this be true?: i think so! expressed uncertainty in online rumoring. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, CHI '16, 360–371. New York, NY, USA: ACM.

Starbird, K. 2011. Digital volunteerism during disaster: crowdsourcing information processing. 7–12.

Stewart, M., and Gail Wilson, B. 2015. The dynamic role of social media during hurricane #Sandy: an introduction of the STREMII model to weather the storm of the crisis lifecycle. 54.

Tausczik, Y.; Faasse, K.; Pennebaker, J. W.; and Petrie, K. J. 2012. Public anxiety and information seeking following the H1N1 outbreak: blogs, newspaper articles, and Wikipedia visits. *Health Communication* 27(2):179–185.

Turner, R. H. 1964. Collective behavior. *Handbook of modern sociology* 382–425.

Vieweg, S.; Hughes, A. L.; Starbird, K.; and Palen, L. 2010. Microblogging during two natural hazards events: what twitter may contribute to situational awareness. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '10, 1079–1088. New York, NY, USA: ACM.

Vlahovic, T. A.; Wang, Y.-C.; Kraut, R. E.; and Levine, J. M. 2014. Support matching and satisfaction in an online breast cancer support community. In *Proceedings of the 32Nd Annual ACM Conference on Human Factors in Computing Systems*, CHI '14, 1625–1634. New York, NY, USA: ACM.

World Health Organization. 2018. World Health Organization disease outbreaks. http://www.searo.who.int/topics/disease_outbreaks/en/, Last accessed on 2018-09-10.

Yoo, S.-W.; Kim, J.; and Lee, Y. 2018. The effect of health beliefs, media perceptions, and communicative behaviors on health behavioral intention: an integrated health campaign model on social media. *Health Communication* 33(1):32–40.

Zhao, Y., and Zhang, J. 2017. Consumer health information seeking in social media: a literature review. *Health Information & Libraries Journal* 34(4):268–283.