UNIVERSITY OF SOUTHAMPTON

Towards an Understanding of Uncertainty on Wikipedia During Crises: A Mixed Methods Approach

by

Reham Al Tamime

A thesis submitted in fulfillment for the
degree of Doctor of Philosophy

in the
FACULTY OF PHYSICAL SCIENCES AND ENGINEERING
Electronics and Computer Science
Web Science Institute

January 2020

This thesis proposes an interdisciplinary approach to understanding the Wikipedia editors' response to new disease outbreak crises. A mixed methods triangulation-based approach is used to examine the change in the editing patterns and the response to the uncertainty during new disease outbreak crises on Wikipedia. Quantitative data sources have been utilised to reveal the patterns of editing activities during new disease outbreak crises and to establish the reasons behind this change. Qualitative data sources have been used to distinguish between the forms and the strategies used to manage the uncertainty during new disease outbreak crises. The triangulation and the integration of results revealed Wikipedia as a social machine. This is further discussed while highlighting the importance of 'co-creation' on Wikipedia during new disease outbreak crises. This thesis presents a thematic framework that depicts various forms of uncertainty and the strategies used to manage this uncertainty on Wikipedia during crises. Therefore, this thesis provides a conceptual and methodological contribution to the field of Web Science and has important implications for future research.

Contents

List of Figures

List of Tables

Declaration of Authorship

I, Reham Al Tamime, declare that this thesis titled, **Towards an Understanding of Uncertainty on Wikipedia During Crises: A Mixed Methods Approach** and the work presented in it are my own. I confirm the following:

- This work was done wholly or mainly while in candidature for a research degree at this University.

- Where any part of this thesis has previously been submitted for a degree or any other qualification at this University or any other institution, this has been clearly stated.

- Where I have consulted the published work of others, this is always clearly attributed.

- Where I have quoted from the work of others, the source is always given. With the exception of such quotations, this thesis is entirely my own work.

- I have acknowledged all the main sources of help.

- Where the thesis is based on work done by myself jointly with others, I have clarified exactly what was done by others and what I contributed myself.

- Parts of this work have been published as follows:

Reham Al Tamime, Wendy Hall, and Richard Giordano. Medical Science in Wikipedia: The Construction of Scientific Knowledge in Open Science Projects. In Proceedings of the 12th International Symposium on Open Collaboration Companion, OpenSym '16, pages 4:1-4:4, New York, NY, USA, 2016. ACM. ISBN 978-1-4503-4481-4.

Reham Al Tamime. Behind the Wikipedia Medical Knowledge Factory: Understanding the Knowledge Dynamic Over Time. In Proceedings of the 2017 CHI Conference Extended Abstracts on Human Factors in Computing Systems, CHI EA '17, pages 238(243),

New York, NY, USA, 2017. ACM. ISBN 978-1-4503-4656-6.

Reham Al Tamime, Richard Giordano, and Wendy Hall. Humans and Bots in Controversial Environments: A Closer Look at their Interactions. Qatar Foundation Annual Research Conference Proceedings, 2018(3):ICTPD425, March 2018a.

Reham Al Tamime, Richard Giordano, and Wendy Hall. Observing Burstiness in Wikipedia Articles during New Disease Outbreaks. In Proceedings of the 10th ACM Conference on Web Science, WebSci '18, pages 117-126, New York, NY, USA, 2018b. ACM. ISBN 978-1-4503-5563-6.

Reham Al Tamime, Richard Giordano, and Wendy Hall. Uncertainty During New Disease Outbreaks in Wikipedia. Proceedings of the International AAAI Conference on Web and Social Media, 113(01):38–46, June 2019. ISSN 2334-0770.

Signed ——-

Date ——

Acknowledgements

For their incredible help over the course of my PhD, I would like to thank my supervisors, Dr Richard Giordano and Professor Wendy Hall. The advice, guidance, and feedback that they have offered me at every stage of my research has been invaluable in leading towards the production of this thesis, and for this, I am incredibly grateful. I would also like to thank my friends and family who have both supported me and endured the many discussions about my work.

Thank you all!

Chapter 1

Introduction: Wikipedia, Uncertainty, and Crises

'I invented the Web just because I needed it, really, because it was so frustrating that it didn't exit.'

— Sir Tim Berners-Lee

With more than 51.2% of the global population, or 3.9 billion people now wired to the World Wide Web (International Telecommunication Union), it is difficult to not notice how far and how fast the Internet usage has rapidly grown since 1990. Driven by Tim Berners-Lee's vision to create a social effect and help people work together (Berners-Lee and Fischetti, 1999), the Web has revolutionised the way people interact, learn, and communicate. Indeed, the Web has created more opportunities for people to communicate with friends and family; access news and information; access services related to e-government, e-health, digital finance, and smart transportation; and enjoy other advantages (International Telecommunication Union, 2018). These advantages could also include accessing and providing critical and rudimentary information during crises such as to register patients, report symptoms, or to allocate aid. Moreover, Web applications such as GeoChat[1] were designed to help team members communicate their location through SMS, email, and Twitter so as to aid faster and more coordinated responses to disease outbreaks and natural disasters (Wazny, 2018).

This thesis approaches the phenomenon of crisis response on the Web by looking at Wikipedia as a major online encyclopedia. In particular, this thesis aims to raise questions about Wikipedia editors' response to crises. The field of Web Science encourages the application of theories, methodologies, and problem-solving approaches from multiple disciplines to address the pressing challenges around the World Wide Web and the

---

[1]https://instedd.org/technologies/geochat

society as a whole. Following a Web Science interdisciplinary approach, this thesis investigates Wikipedia as a socio-technical phenomenon. As such, this thesis implements a mixed methods approach and integrates multiple disciplines including computer science, health sciences, and social sciences. As the first step, this chapter provides an introduction to Wikipedia and social machines and defines uncertainty and crises as both are important concepts related to this research. In addition, the interconnection between Wikipedia and these concepts are presented in this chapter with an outline of the scope of this thesis.

1.1 Wikipedia

Wikipedia is the most influential popular information source on the Internet and is ranked as the fifth most visited website[2]. Founded in 2001, Wikipedia attracted more than 16.14 billion total monthly page views as of December 2018[3]. There are approximately 72,000 active contributors working on more than 48,000,000 articles in 302 languages[4]. Wikipedia permits anyone to contribute to articles by simply clicking on the 'edit' button. Every revision is stored in Wikipedia sequentially to enable users to view, modify, or revert to (restore) the previous revisions. Figure 1.1 shows the revision history page of the Zika Virus Wikipedia article. In addition, the built-in talk pages serve as a space for discussions about the articles' content. Figure 1.2 shows the talk page of the Zika Virus Wikipedia article. Even though Wikipedia is self-managed by editors who adhere to policies and guidelines[5], editors can appoint administrators who become involved in specialised tasks, such as reviewing articles at others' requests, tagging articles as controversial, and remaining vigilant for vandalism. Similar roles are undertaken by Wikipedia bots[6], which are built to automate certain tasks, such as correcting typographical errors, detecting vandalism, and modifying references. Tables 1.1 and 1.2 provide an overview of the Wikipedia page views statistics as of December 2018.

Total page views	16.14 billion
Total edits	43.08 million
New registered users	363.14 thousand

Table 1.1: Wikipedia Monthly Statistics

Between 2001 and 2007, Wikipedia witnessed exponential growth in both the number of articles and the number of users that were contributing to new articles (Almeida et al., March 2007). This growth, however, did not sustain as the number of editors and the

---

[2]https://www.alexa.com/topsites

[3]https://stats.wikimedia.org/

[4]https://en.wikipedia.org/wiki/Wikipedia:About

[5]https://en.wikipedia.org/wiki/Wikipedia:Policies_and_guidelines

[6]https://en.wikipedia.org/wiki/Wikipedia:Bots

USA	3.41 billion
Japan	1.07 billion
Germany	1.07 billion
United Kingdom	838.72 million

TABLE 1.2: Wikipedia Monthly Page Views by Country

number of edits per month started to decline at the beginning of 2007 (Suh et al., 2009). This decline has been attributed to many factors such as the increasing completion of articles in the context of a population model as well as the decreasing newcomers' retention (Halfaker et al., 2013). Nevertheless, the assumption that Wikipedia has entered a period of decline has been challenged by Tinati et al.. By using WikiProjects[7] as a proxy for examining the contribution and discussion activities, the researchers argued that Wikipedia is still enjoying a period of growth of new users, edits, and discussion activities (Tinati and Luczak-Roesch, 2017).
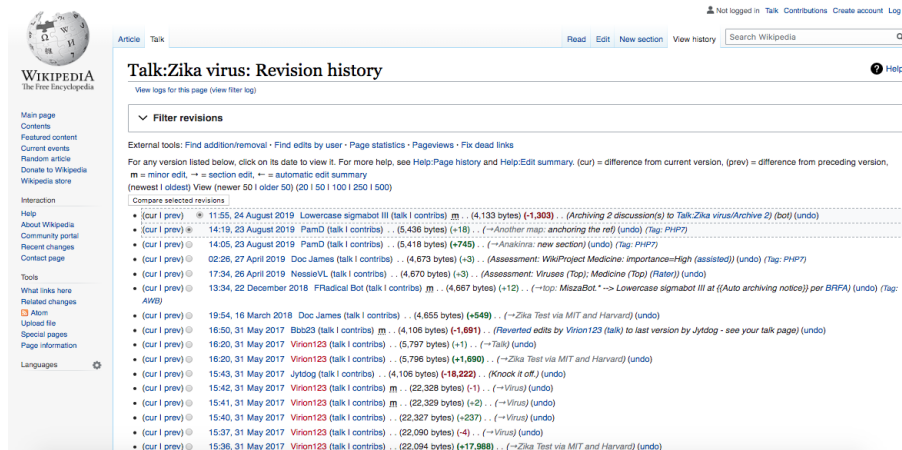
FIGURE 1.1: Wikipedia Revision History
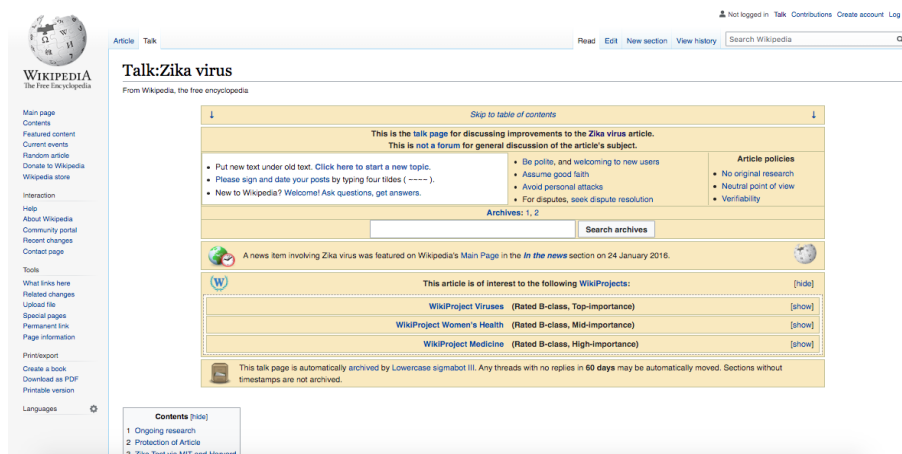
FIGURE 1.2: Wikipedia Talk Pages

---

[7]https://en.wikipedia.org/wiki/Wikipedia:WikiProject

1.2   Wikipedia as a Social Machine

The concept of social machines was first introduced by Berners-Lee and Fischetti, as "real life is and must be full of all types of social constraints – the very processes from which society arises. Computers can help if we use them to create abstract social machines on the web: processes in which the people do the creative work and the machine does the administration" (Berners-Lee and Fischetti, 1999). This definition of social machines recognises the joint contribution of humans and machines to a particular process on the Web. Nevertheless, this definition does not acknowledge the machines that could do both administrative and creative work. Smart et al. (2014) challenged Berners-Lee and Fischetti's definition by looking at systems such as Picbreeder. Picbreeder is a system that supports the collaborative and interactive production of images by using a mixture of evolutionary computation techniques and human agents. The machine in this particular example plays a creative role, which is to generate and publish images after humans select their features. This example shows that it became necessary to prevent restricting and assigning certain types of actions, such as administrative or creative actions, to either humans or machines. Therefore, social machines have been defined as "Web-based socio-technical systems in which the human and technological elements play the role of participant machinery with respect to the mechanistic realization of system-level processes" (Smart et al., 2014). This definition implies that social machines have several properties that need to be highlighted. First, social machines involve human and technological components working together. Second, social machines are Web-based systems. Third, social machines consist of multiple individuals and can exploit differences between these individuals. Fourth, social machines are physical systems that perform, implement, or realise processes (Smart et al., 2014).

Wikipedia is a social machine that is co-created by human participants and technological components (Shadbolt et al., 2019). This term 'co-creation' is important to highlight, as there is no dividing line between functions that are performed by human participants and functions that are performed by technological components. In fact, in social machines, it is inappropriate to refer to participants as 'users' as they do not simply use a given technology, but they engage and drive interactions to solve problems (Shadbolt et al., 2019). As such, systems in which humans solely browse or consume content without contributing to processes and activities in return are not classified as social machines (Smart and Shadbolt, 2014).

1.3   Uncertainty

Even though uncertainty is embedded in every aspect of our life, a uniform definition of uncertainty does not exist. Nevertheless, uncertainty is usually defined as a lack

of predictability regarding a situation, individual, or behaviour (Berger and Calabrese, 1975). Further work elaborated that "uncertainty exists when the 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). At the same time, other researchers emphasised 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).

Uncertainty can occur as a result of information deficiency. This deficiency could be in the form of incomplete, fragmented, contradictory, unreliable, or vague information (Klir, 2005). Uncertainty is also seen as the non-specificity, discord/strife, and fuzziness of information (McCauley-Bell and Freeman, 1997). The non-specificity of information is defined as unspecific sets of alternative choices. This type of uncertainty can arise as a result of having a variety of choices, generality, equivocation, and imprecision in information. Moreover, the discord/strife of information is defined as a disagreement or a discrepancy among a set of alternative options. This type of uncertainty can emerge as a result of conflict, incongruity, and dissonance. Comparably, fuzziness is defined as a lack of a definite or sharp distinction among a set of alternatives. This type of uncertainty becomes noticeable when there is either a vague, an unclear, or an obscure set of alternatives (McCauley-Bell and Freeman, 1997). Others view uncertainty as states that have no institutional knowledge such as legal systems and standards to facilitate coordination or scientific evidence (Aspers, 2018). This shows that institutional knowledge can play a role in minimising uncertainty.

1.4 Uncertainty, Crises, and Disease Outbreaks

Uncertainty is particularly inherent in the field of medicine and arises in every step related to the diagnosis, prognosis, and treatment of diseases. The medical source of uncertainty is experienced not only in the cases of chronic diseases (Miller, 2012; Mishel, 1999), but also in the cases of new medical problems that can stimulate a source of uncertainty that did not previously exist (Albrecht et al., 2000). Uncertainty can cause serious concern and could adversely affect physicians and patients simultaneously, as it complicates the ability to prevent, diagnose, and treat an illness. Therefore, the sources of uncertainty around medicine and public health are usually amenable to research and analysis (Albrecht et al., 2000).

There is a distinct type of uncertainty that is also important to analyse, which is the uncertainty that is experienced during new disease outbreaks. Disease outbreaks are crises that are characterised by "high consequentiality, limited time, high political salience,

uncertainty, and ambiguity" (Moynihan, 2008). In other words, disease outbreaks are unforeseen events that have a severe societal impact and require close and immediate interactions among the affected organisations, particularly among public authorities (Müller-Seitz, 2014). This impact can occur at the national, local, and individual levels. For example, individuals had to re-consider and probably adjust their travel plans to different Zika-affected countries during the disease outbreak.

1.5   Tying It All Together: Wikipedia, Uncertainty, Crises, and Disease Outbreaks

Wikipedia plays a major role during crises and emergency situations. Evidence has shown that the number of visits to certain Wikipedia articles increases immediately after crises or major incidents. For instance, the number of visits to the H1N1-relevant Wikipedia articles peaked shortly after the World Health Organization's announcement of the H1N1 disease outbreak (Tausczik et al., 2012). Moreover, links to Wikipedia articles that are related to recent crises could be visited and circulated around different social media channels such as Twitter (Heverin and Zach, 2010). Beyond the number of visits, editing activities to the corresponding articles also increase immediately after crises or catastrophes (Keegan et al., 2011). For example, the authorship traffic grew quickly as a result of the contribution of new information on Wikipedia after the Virginia Tech crisis (Palen et al., 2007). This is particularly significant as crises are often characterised by uncertain and constantly changing information, which, in turn, demands Wikipedia editors to retrieve and understand information that is often complex, ambiguous, and contradictory. In this respect, this makes an important case study to further understand the editing patterns and the response to uncertainty on Wikipedia during major crises, particularly during disease outbreaks.

1.6   Outline of This Thesis

This chapter presents the fundamental concepts related to understanding the Wikipedia editors' response to crises. Consequently, this thesis aims to investigate 1) the change in editing activities on Wikipedia during crises; 2) the factors that cause changes in editing activities on Wikipedia during crises; 3) the forms of uncertainty that are expressed by Wikipedia editors during crises; and 4) the strategies that are used by Wikipedia editors to manage the uncertainty during crises. With a focus on new disease outbreaks, the findings will be used to build a framework that describes the uncertainty around the Wikipedia editors' response to crises. Ultimately, this thesis presents a conceptual and methodological contribution to the field of Web Science. This section outlines the research steps and structure that were used to make this contribution.

Chapter 2 provides an in-depth introduction to Wikipedia by looking at previous studies including but not limited to the dynamics of viewing and editing content, editors' characteristics, and controversies on Wikipedia. Chapter 3 then moves to an assessment of the relevant literature related to crises, crisis response, and health. This assessment covers not only the studies on Wikipedia but also the studies on social media and the Web in general. These two chapters provide the foundation to find the gap in knowledge and, hence, to set up the research questions for this thesis.

Chapter 4 introduces the methodological approach that was adopted to answer the research questions. In order to provide interdisciplinary insights into the problem, case study design was selected in conjunction with a mixed methods pragmatism approach. Chapter 5 follows up by introducing the chosen case study and explaining the data collection, methods, and the qualitative and quantitative analysis techniques. While Chapter 4 lays the procedural framework to carry out the research and help to select the research methods in general, Chapter 5 describes the selected research methods including the case study, data sources, and qualitative and quantitative techniques and tools in specific.

This thesis then progresses to reporting the quantitative results in Chapter 6. This chapter aims to explore and compare editing activities between articles related to new disease outbreaks and articles that are not related to new disease outbreaks (articles that are related to chronic diseases). Chapter 7 further elaborates on the quantitative findings by focusing on studying different factors that can explain the results obtained by analysing new disease outbreak articles. These two chapters help to distinguish among editing practices around new disease outbreak crises on Wikipedia.

Moreover, Chapter 8 reports the qualitative findings as derived from Wikipedia articles that are related to new disease outbreaks and chronic diseases. These findings reveal various forms of uncertainty as well as strategies that were implemented by Wikipedia editors to cope with uncertainty. The qualitative results are then used to construct a framework that describes the uncertainty during crises on Wikipedia. This framework is further refined in Chapter 9.

In Chapter 9, the qualitative and the quantitative data are discussed and then triangulated in relationship to the research questions, literature, and existing theories. Accordingly, a revised version of the framework is presented to lead to the overall conclusion of this thesis. Finally, the limitations of this work and the potential for further study are examined closely, before presenting the final remarks to summarise this thesis and outline the overall contribution.

Chapter 2

Background: Current Understanding of Wikipedia

'Imagine a world in which every single person on the planet is given free access to the sum of all human knowledge. That's what we're doing.'

— Jimmy Wales

This chapter forms a necessary base for understanding Wikipedia as a collaboratively created and edited encyclopedia on the Web. Fundamentally, this chapter includes studies on the dynamics related to viewing and editing content on Wikipedia since its establishment in 2001. This also includes looking closely at various characteristics related to the human and the bots' participation in the creation of Wikipedia articles. Participation in content creation can involve controversies and disagreements between editors. Therefore, this chapter presents studies related to conceptualising and measuring controversies on Wikipedia. As a starting point, this chapter provides insights about Wikipedia as both a social and a technical phenomenon that is important to study.

2.1 Wikipedia Page Views

Considering that Wikipedia has a huge impact on how people access and gather encyclopaedic information online, several researchers have attempted to gain a better understanding of the different aspects related to accessing information on Wikipedia. Primarily, page view statistics[1], i.e. the number of visits to an article, have been used to unmask the patterns and dynamics related to viewing content on Wikipedia. For example, Singer et al. (2017) combined survey responses and Wikipedia page views to capture the users' motivations to visit Wikipedia articles under different categories. The study recognised that Wikipedia might be used for both shallow and deep information

---

[1]https://en.wikipedia.org/wiki/Wikipedia:Pageview_statistics

needs. Shallow information needs often involve fact checking or randomly exploring certain topics such as topics related to sports. Deep information needs often involve in-depth navigation through certain articles such as through articles that are related to the sciences and the arts. Deep information needs are usually driven by intrinsic learning, while shallow information needs are driven by the need for making a personal decision, by the media, or simply by the interest of bored/random exploration. Other studies looked at predicting the popularity of an article during the first hour of exposure by using the hourly and the daily number of page views (Thij et al., 2012). In addition, page views have been used to find pages of interest and cluster these pages into coherent topics (Ahn et al., 2011).

An increase in the Wikipedia page views for specific articles may reflect social trends and the interest in events occurring worldwide (Khan et al., 2015). In fact, several studies have demonstrated that Wikipedia page views can be used to measure and identify the trends related to real-world events (Tinati et al., 2013). For instance, studies focused on the possibility of using page views statistics during electoral campaigns to monitor and predict the popularity of candidates and parties (Yasseri and Bright, 2013), and the overall vote share of political parties (Yasseri and Bright, 2016). Moreover, a subsequent study relied on page views to investigate the impact of scientific awards on the information about scientists and their research topics on Wikipedia (Wagner et al., 2018). Similar studies found that Wikipedia page views can successfully predict the box office revenue for films released in the United States during their opening weekends in theatres (de Silva and Compton, 2014; Mestyán et al., 2013). Studies also looked at the potential of using page views for economic and financial applications such as to uncover stock prices under different market regimes (Cergol and Omladič, 2015), detect the early signs of stock market movement (Liu et al., 2014; Moat et al., 2013, 2014), and provide investors with insights about the performance of a company's stock on the basis of how often the company's Wikipedia page is viewed (Wei and Wang, 2016).

Various tools and approaches have been developed on the basis of the Wikipedia page views statistics to monitor and predict the trending topics and events. These approaches include using a sliding window implementation to forecast the trending topics (Holaker and Emanuelsen, 2013), and using the nearest neighbour approach to exploit the semantic relationships between topics (Althoff et al., 2013). On the basis of page views, researchers have also considered defining contextual measures, including semantic concepts and semantic relationships between pages in different languages, in order to predict the emerging topics (Kämpf et al., 2015). Other researchers have developed tools that allow the selection of the topics of interest to find the popularity of the related articles on the basis of the corresponding page views (Ciglan and Nirvåg, 2010).

2.2   Wikipedia Page Edits

Research has been extended to better understand the dynamics of participation and the editing activities on Wikipedia. For instance, Kuznetsov (2006) identified five values that underline motivations to contribute to Wikipedia. These values include reputation, community, reciprocity, altruism, and autonomy. A similar study agreed that Wikipedia contributors have mainly pro-social motivations such as a desire to give back and make a difference in the lives of others (Antin, 2011). However, research has also suggested that the contributors' motivation changes over time as non-instrumental motivations such as collective, fun, and self-expression can significantly decrease over time (Balestra et al., 2017).

Thus far, researchers have focused on theoretically explaining the contributors' interactions and knowledge-building activities. Cress and Kimmerle (2008) used Luhmann's systems theory and Piaget's cognitive theory to understand the collaborative knowledge building in wiki systems. Luhmann differentiated between two types of systems: cognitive and social. Cognitive systems operate via consciousness and cognitive processes, such as through the retrieval of knowledge from long-term memory, whereas social systems operate through a means of communication. Luhmann's theoretical model realised that the consciousness of wiki contributors enables the emergence of cognitive processes, while the Wikipedia systems and infrastructure enable the emergence of social processes. Thus, collaborative knowledge building is achieved by coupling the cognitive and the social processes. Coupling the social and the cognitive processes involves introducing, communicating, receiving, and integrating knowledge. In particular, cognitive processes can be seen in wiki systems as individuals need to acquire, understand, and learn new knowledge in order to edit and update Wikipedia articles. Furthermore, social processes can be seen in wiki systems as individuals need to share and cooperate to create and update Wikipedia articles. Piaget's cognitive theory (1976) was developed primarily to study the cognitive development of children and adolescents (Piaget, 1976). Focusing on the cognitive processes, Piaget's cognitive theory was incorporated to better understand the collaborative knowledge building on wiki systems. This theory describes the way individuals try to maintain a balance between the environmental information on the one hand and their prior knowledge on the other hand. The theory is useful when attempting to understand the cognitive development of contributors in wiki systems such as through the process of assimilating and accommodating knowledge. In particular, wiki contributors assimilate knowledge by understanding new knowledge and then integrating this knowledge into their existing knowledge. However, wiki contributors accommodate knowledge by interacting with information to not only assimilate and integrate knowledge, but also change their existing knowledge. Swarts (2009) selected Latour's Actor Network Theory (ANT) to understand mass collaborative knowledge creation. ANT

helps to describe the construction of content in Wikipedia as a series of connections between human and non-human actors, such as authors, editors, information, references, and interfaces. These actors interact with one another to form a network that often leads to stable Wikipedia entries or facts, as the author suggested. Following this theoretical perspective, the author selected the clean coal technology Wikipedia article as a case study. A qualitative analysis of the editors' discussion was performed by coding and tracking the iterations of statements until they become accepted as facts. This analysis revealed that the editors used different strategies to form stable Wikipedia entries, including pushing facts forward, creating alliances, and hiding controversies.

Researchers have also noticed that the dynamics of several editing activities reflects the attention given to specific articles after major events. Real-world events related to political conflicts, natural catastrophes, and new scientific findings can trigger a surge in the number of edits in the corresponding Wikipedia articles (Georgescu et al., 2013). For example, Wikipedia articles related to *Black Lives Matter* experienced a period of high activity after the Black Lives Matter social movement. In addition, the number of edits was found to correlate with the number of page views in specific articles related to the event. This showed that major events can reshape editing activities as well as information seeking behaviours on Wikipedia (Twyman et al., 2017). Other work showed that Wikipedia's biographies are intensively updated following the death of their subject (Keegan and Brubaker, 2015). Zhang et al. (2019) studied the collaborative dynamics of Wikipedia articles after times of shock. Focusing on articles about people, the experiments revealed that participation during times of shock is different from participation during normal times, as both the newcomers and the incumbents tend to participate at higher rates during shocks.

The number of edits to Wikipedia articles is triggered not only by external factors such as real-world events, but also by internal factors. For instance, editors are likely to be more active in editing articles before awarding or receiving barnstars.[2] (Lim et al., 2013). In addition, Das et al. (2016) found that a large number of editors change their behaviours and begin focusing more on a particular controversial topic once they are promoted to the administrator status. While Zhang et al. (2017) confirmed that editors change their behaviours before articles' evaluation periods. In particular, editors might increase their level of activity and decrease conflict (i.e. reverts) prior to the articles' nomination. Nevertheless, editors revert to their typical editing behaviours after the promotion of an article.

---

[2]which is an informal reward system where editors award barnstars to other editors in recognition of their contributions to Wikipedia.

2.2.1 Wikipedia Editors' Characteristics

Researchers have also paid attention to understanding various characteristics related to Wikipedia contributors such as the gender and type of the contributors (i.e. humans and bots). For example, studies have found a gender gap in Wikipedia as the number of male editors is larger than the number of female editors. Despite these findings, the study recognised that the most active women (i.e. as measured by ranking the number of revisions) tend to make larger revisions than the most active men (Antin et al., 2011). Other work reported that female administrators and non-administrators behave similarly as they tend to often cite Wikipedia policies (Laniado et al., 2012).

Numerous researchers have recognised the increasing significance of bots in both the editorial and the administrative processes (Geiger, 2009; Geiger and Ribes, 2010). In addition, Geiger and Halfaker (2013) found that during ClueBot NG's downtime, the median time-to-revert had nearly doubled. Further, Müller-Birn et al. (2013) analysed the growing relevance of algorithmic governance in the realm of governing large-scale peer production communities. Other studies have reported the contribution of bots not only to the English Wikipedia, but also at a global level and across several languages (Steiner, 2014; Niederer and van Dijck, 2010). Some others have studied different social phenomena such as the growth of Wikipedia and the response to unexpected events by incorporating both humans and bots as members of the Wikipedia community (Tinati and Luczak-Roesch, 2017). Different researchers have extended their scope to look not only at the edit reverts between humans in Wikipedia, but also at the edit reverts between bots (Tsvetkova et al., 2017). This work found that bots tend to revert each other's edits or *fight* for years. These fights mostly occur between bots that specialise in creating and modifying links between editions of different languages of the encyclopedia. This, however, could be attributed to a lack of coordination and the differences in the rules and the conventions across different language editions.

2.3 Wikipedia and Controversies

As discussed in the previous section, editing Wikipedia articles can involve both cognitive and social processes. Social processes suggest that editors need to communicate in order to update and edit Wikipedia articles. As a result, different types of communication can emerge between editors such as agreement, negotiation, disagreement, or even controversies. Controversies are defined as "situations where actors disagree" (Venturini, 2010). The term *controversy* refers to a specific type of exchange that involves a difference in opinion that might be difficult to resolve (Van Eemeren and Garssen, 2008). This term describes a shared uncertainty. Controversies begin when different actors cannot ignore each other and end when actors manage to work out a solid compromise and

reach an agreement (Venturini, 2010). Further, a controversy is a type of debate that starts with a specific problem and gets escalated to other problems to reveal profound divergences. Divergences can include opposing attitudes and preferences (Van Eemeren and Garssen, 2008). Controversies can be escalated to become conflicts if they involve intense disagreement and violence. Nevertheless, a controversy is a complicated social phenomenon that is often difficult to formally and systematically define (Jang et al., 2017).

Wikipedia is increasingly becoming a "rich historical source logging the development of societal controversies over time" (Borra et al., 2015). Controversies in Wikipedia are unavoidable as a consequence of its open nature, which means that it accepts contradictory views on a topic. Consensus on Wikipedia can be achieved through "user-run quasi-formal mediation and arbitration rather than through mechanical control of behaviour" (Berry, 2012). Different features such as page edit history and talk pages encourage editors to discuss and resolve controversies on Wikipedia. This has enabled researchers to measure and quantify the controversy on Wikipedia.

Kittur et al. (2007) measured controversy by counting the number of 'controversial' tags in the history of an article. Suh et al. (2007) focused on detecting controversies and conflicts in Wikipedia by building a model that uses the number of reverted edits as an indicator of the controversies within an article. In addition, the level of dispute has been used as a metric to identify controversial articles. The level of dispute is measured by the number of deleted words when two authors revise each other's work (Vuong et al., 2008). Bogdanov et al. (2010) proposed a topic modelling approach to identify a group of contributors who share views on certain topics and collectively oppose other groups. Furthermore, Maniu et al. (2011) explored Wikipedia editors' interactions, such as reverting or deleting content, as well as voting, as indicators of positive or negative relationships between editors. Bykau et al. (2015) came up with a novel approach to detect the controversies in Wikipedia. This approach focuses on computing the amount of content substitution. The approach assumes that authors who have opposing viewpoints tend to substitute each other's content, rather than just inserting or deleting content. Sepehri-Rad and Barbosa (2015) used Wikipedia's administrative elections and voting data to construct a machine learning formula that predicts the positive (supporting) and negative (opposing) attitudes of editors. Rad and Barbosa (2012) conducted different experiments to model the controversy in Wikipedia. These models can be classified into four groups: meta-driven, user-driven, content-driven, and pattern-driven. The meta-driven approach relies on extracting simple statistics from the revision history and the discussion pages such as the number of revisions and the number of unique editors. The user-driven approach relies on measuring the agreement and the disagreement between editors on the basis of their interactions such as on the basis of the number of words restored and deleted from a reverted edit. However, the content-driven approach depends

on analysing content in the revision history of an article to track the authorship and the deleted words. Finally, the pattern-driven approach is based on analysing the patterns of edits over a history of revisions so as to identify the mutual reverts as a sign of an edit war. Sumi et al. (2011) and Yasseri et al. (2012) claimed that none of the previous measures of controversy are sufficiently accurate to catch all the controversial instances within an article. Therefore, they developed a model that ranks and detects controversial articles on the basis of the reverted edits between a pair of editors. This model has been applied to analyse the temporal characteristics of editing activities and behaviours in controversial and non-controversial articles based on macro (years) and micro (hours, days, and weeks) timescales. Furthermore, the model has been used to find and study controversial articles across different languages and cultures (Yasseri et al., 2013).

Visualization approaches have been developed to surface editors' interactions, controversies, and conflicts over time. Viégas et al. (2004) introduced 'history flow', which is a tool that unmasks the editors' social behaviours such as negotiation and consensus. Brandes and Lerner (2008) constructed a visualisation tool that reveals the most dominant authors of a page, the roles that they play, and their confrontations, on the basis of their edit history and interactions. This was achieved by studying the editors' reverted edits as an indicator of disagreement. Ekstrand and Riedl (2009) reported that the linear representation of the revisions' history in Wikipedia hides the relationships between revisions. Therefore, they used the linear representation of the history view to create relationships between revisions and detect the discarded work, which serves as a tool for understanding the conflict between the editors. Focusing on the feminism articles in Wikipedia, Berry (2012) used the available visualisation techniques such as Edit Scrapper and Dorling Map to map the concerns and the controversies related to the issue of feminism over time. This work suggested that it is difficult to trust or mistrust Wikipedia articles without seeing how the disputed content unfolds over time. Flöck and Acosta (2015) proposed WhoVIS, a web tool to mine and visualise the editors' interactions in Wikipedia over time. The authors defined specific types of interactions between editors: delete a number of words, undo the deletion of a number of words, and reintroduce a number of words (if a number of words were deleted by an editor and then reintroduced again by another editor in a later revision). These interactions were used to compute and visualise the disagreement between editors. WhoVIS helped to reveal the patterns of disagreement over time that would otherwise be difficult or impossible to obtain from the existing tools.

2.4   Summary

Web Science research requires an understanding of both the social and the technical aspects of the Web (Tim Berners-Lee, 2007). Accordingly, this chapter introduced

Wikipedia as a socio-technical system. This was done by discussing that not only is Wikipedia supported by a technical infrastructure, but it also supports and hosts several cognitive and social processes. By examining the literature related to the Wikipedia page views and the Wikipedia page edits, we found that Wikipedia's history logs of page views and page edits are not merely technical features; these features also reflect a variety of global trends and societal interests. Scholars have acknowledged that humans (both male and female) and bots play a role in editing Wikipedia articles. Moreover, the literature related to Wikipedia page edits also emphasised that the interaction and the communication between editors are as important as finding and understanding information in order to create and update content on Wikipedia articles. Interactions between editors may involve disagreements and controversies; hence, this chapter suggested that research on studying and measuring the controversy on Wikipedia is still ongoing. By concluding that studies on Wikipedia recognise the importance of the social activities and interactions within, this chapter affirmed that Wikipedia could be used to study social phenomena such as a crisis response. The following chapter will focus on the literature related to the crisis response on Wikipedia and the Web in general.

Chapter 3

Background: Crises and Crisis Response on the Web and Wikipedia

'We cannot stop natural disasters but we can arm ourselves with knowledge: so many lives wouldn't have to be lost if there was enough disaster preparedness.'

— Petra Nemcova

The previous chapter argued that it is important to study Wikipedia as a socio-technical phenomenon. Building on this understanding, this chapter focuses on reviewing the literature related to Wikipedia in the context of crises. This chapter starts by providing background information about crises and crisis response on the Web in general. Then, this chapter advances to review the literature on crisis response on Wikipedia in particular. As this thesis selects new disease outbreaks as a type of crises to investigate, this chapter does not ignore examining the literature related to health and well-being on Wikipedia. This review helps to identify research gaps and leads to the formulation of the research questions for this thesis.

3.1 Crises and the Web

Crises are conditions of high uncertainty and threats that need swift responses and decision making. The terms *emergency*, *crisis*, *disaster*, and *catastrophe* are used interchangeably to describe the scale and significance of such situations (Eriksson and McConnell, 2011). During crises, the use of social media increases. Overall, people use social media during crises for various reasons, such as to seek neighborhood-specific interactive forums; 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 et al., 2012). Moreover, particular social media channels are perceived to be more credible than traditional mass media during crises (Austin et al., 2012).

The prevalent use of social media during crises and emergencies has led researchers to explore the characteristics of online information generated during various crises. For example, Vieweg et al. (2010) examined the situational features that were communicated on Twitter during two concurrent emergency events that took place in the US: the Red River Floods and the Oklahoma Grassfires. The study identified different types of information related to situational updates on the crises. These included geo-location and location-referencing information such as warning, preparatory activity, fire line or hazard location, flood level, and weather, wind, visibility, and road conditions. A further analysis of twitter data during the 2009 Red River Floods disclosed that retweeting information acts as an informal recommendation service, as locals are more likely to use the retweet convention to pass on emergency-related information during the event (Palen et al., 2010). Other studies found that people during the 2008 Sichuan Earthquake preferred to reply to threads on the Tianya forum that aimed to gather and integrate information. However, they also tended not to reply to threads that described the author's participation in disaster relief or provided emotional support (Qu et al., 2009). These studies collectively reflect the significance of social media use during crises for information and emotional support.

Different researchers studied the perceptions of uncertainty and risks on social media during crises and disease outbreaks. In particular, Starbird et al. (2014) looked at three false rumours that were circulated on Twitter in the aftermath of the 2013 Boston Marathon Bombing. Although the researchers found evidences of crowd-correction for each rumour, these corrections were considerably smaller than the propagation of misinformation. In addition, Starbird et al. (2016) sought to understand how uncertainty was expressed on Twitter across the lifecycle of different rumours associated with the Boston Marathon Bombings and Sydney Siege events. The study identified several milling behaviours in crisis-related rumours that disclose the process in which people try to make sense and find possible explanations for uncertain events. These behaviours include interpreting, speculating, theorizing, debating, and challenging specific rumours. Other work described a qualitative analysis on people's discussions on Reddit, BabyCenter, and TripAdvisor in order to understand how they made travel decisions during the Zika virus crisis. The researchers observed that people faced extreme uncertainty during the Zika virus crisis because of the limited amount of authoritative and scientific knowledge. This extreme uncertainty led people to rely on local information that was shared by previous travellers or locals. Moreover, this extreme uncertainty led people to infer the probability of Zika in a particular country on the basis of the knowledge about its neighbouring countries (Gui et al., 2017). Furthermore, Gui et al. (2018) explored several

public perceptions of risk on Reddit during the Zika crisis. These perceptions included the interpretations, opinions, and judgments of risks. The study distinguished between two types of risks related to the emerging Zika epidemic: risks caused by the Zika virus such as the risks associated with getting infected, infected babies, and the affected population in general; and risks that might be introduced by disease management measures such as official policies and actions to combat Zika.

3.2 Crises Response and the Web

Disasters and crises create complex conditions that require intra-organisational and inter-organisational coordination to assess and respond to the humanitarian needs of the affected population (Noori et al., 2016). Responding to a crisis usually involves efforts to find, evacuate, search, and rescue victims (Dugdale et al., 2012). Social media applications including social networking sites, crowdsourcing applications, web-based tools, and mobile applications play an important role in disaster and crisis events. Government institutions, international organisations, and non-governmental organisations (NGOs) increasingly use social media as an integral part of crisis management operations in the prevention, preparation, response, and recovery phases. In fact, social media also enable citizens to get involved in crises management operations by the provision, seeking, and brokering of information, by connecting those within and outside the event's geographical space, and by communicating formally and informally with the response effort (Akhgar et al., 2013).

The use of social media applications during crises has proved to have several advantages. For example, social media enable public health officials to quickly and efficiently disseminate information to citizens and accurately pinpoint recovery and rescue efforts. Moreover, the applications of social media in public health surveillance during a disaster may allow the early detection of potential disease outbreaks and the better tracking of disease trends (Finch et al., 2016). Social media facilitate the sharing and accessing of data to understand how sentiments or rumours spread, to engage with the appropriate public health messages, and to deploy the right interventions (Kass-Hout and Alhinnawi, 2013). Considering that different social media applications have been found to contribute to the improvement of communication prior to and during crisis events, different tools have been developed to benefit from the information on social media during crises (Maresh-Fuehrer and Smith, 2016). For instance, the Ushahidi Platform is an open-source web application that was developed to collect, map, and visualise information during crises. Ushahidi allows people to collect and share their own stories using various media such as SMS, web forms, email, or Twitter (Dugdale et al., 2012). Other social media mapping applications including SituMap and PhotoSorter were advanced to cultivate participation, collaboration, and conversation for crisis communication. This is

by including several features that can identify and map locations in real-time, coordinate efforts with other emergency responders prior to and during a crisis, and identify traffic and weather issues to plan emergency response efforts (Maresh-Fuehrer and Smith, 2016).

Several studies have focused on the application of crowdsourcing to effectively and efficiently respond to disasters and crises. Crowdsourcing is the use of ICT to mobilize online volunteers who willingly help the affected people, report crisis situations, translate reported messages, carry out crisis mapping, and self-organize the coordination of relief resources (Park and Johnston, 2015). Social media play a role in mobilising volunteers during crises as social media enable open requests for volunteers to quickly disseminate across the globe (Park and Johnston, 2015). Starbird (2011) suggested using Twitter to crowdsource information during disasters and mass emergency events. This could be achieved by asking Twitter users to incorporate special hashtags into their crisis-related tweets to make these tweets machine-readable and to be able to make sense of the Twitter data in real-time. Other studies implemented a prototype of City-Share, which aims to support the affected citizens, spontaneous volunteers, and public authorities and emergency services. This is by providing a communication infrastructure that encompasses the situated crowdsourcing mechanisms for managing the offers and demands of activities during disasters (Ludwig et al., 2017). Other work applied a mixed methods approach to explore the work practices and interactions of Open Street Map volunteers during disasters. This research succeeded in identifying seven distinct mapping practices that could be classified according to the dimensions of time, space, and interpersonal interactions (Kogan et al., 2016).

There are various bright and dark sides of using different social media applications for crisis management. Stewart and Gail Wilson (2015) indicated that the bright side of using social media during the Hurricane Sandy crisis includes surveillance, social listening, and engaging with the audience (Stewart and Gail Wilson, 2015). In contrast, the dark side of using social media during the Hurricane Sandy crisis includes the spread of rumours, hoaxes, and false and unsubstantiated information. Other studies assessed different uncertainty management strategies that were deployed on Twitter by the Center for Disease Control and Prevention (CDC) during the Ebola outbreak (Dalrymple et al., 2016). The study found that the CDC used Twitter to promote expertise and build trust by communicating the organisation's competence, and stressing the value of following the established procedures and protocols. However, the study also discussed that the CDC found it difficult to initiate a dialogue or engage with the public on Twitter. Instead, the CDC relied on top-down communication in which experts offered advice and answered questions that were mainly related to the risk factors for diseases.

3.3   Crises Response and Wikipedia

As shown in the previous chapter, the Wikipedia page views could be used to measure and predict specific trends and interests in various social phenomena. In the context of crises, the number of views to particular Wikipedia articles was used to assess the anxiety and information-seeking behaviours in response to the 2009 H1N1 outbreak (Tausczik et al., 2012). The number of visits increased after the announcement of the H1N1 outbreak and then declined rapidly. Accordingly, the number of views to particular Wikipedia articles was used to build models and to monitor and forecast the incidence of some diseases. This was done to develop a method that could support traditional disease surveillance tools. McIver and Brownstein (2014) developed a Poisson model that tracks the level of the influenza-like illness in the United States on the basis of Wikipedia daily visits to related articles. Further, the number of daily visits to Wikipedia articles was used by Generous et al. (2014) to build a model that can monitor and forecast the incidence of infectious diseases. The model was successful in monitoring and forecasting the progress of an infectious disease over the course of up to 28 days. However, the method has several limitations such as it needs to be tested using a large number of articles and many more contexts. Moreover, 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 simulations but also needs mathematical theories to describe the flow of disease-related information from human observations through internet systems and algorithms to actionable knowledge.

Different studies have examined how the Wikipedia community responds to unexpected events and emergencies. Keegan et al. (2011) identified several dynamic features of collaboration during unexpected events. These features include intense editing activities concerning articles related to these events. In addition, Wikipedia templates have been created to collect information about events (such as casualty rates) to ensure accurate, and up-to-date information is provided consistently across articles. Further study revealed that articles related to breaking events such as disasters, catastrophes, and conflict exhibit a network structure that is distinct from the structure of the articles on non-breaking events. In particular, articles created immediately after an incident usually tend to have denser networks (Keegan et al., 2012). Additionally, articles related to breaking events exhibit a large interrelated collaboration structure immediately following their creation, whereas non-breaking articles take a year or more to produce a similar structure (Keegan et al., 2012). These findings highlight the role that unexpected events play in motivating participation in online communities. Different researchers have tried to get a clear picture of the psychological processes underlying the users' collaboration while editing Wikipedia articles related to natural and human-made disasters. This

involves a better understanding of the editors' internal thoughts, emotions, and sense-making activities including different ways in which the users organise the thoughts and memories about facts and events. After conducting content analyses on articles covering natural and human-made disasters, the study revealed that articles about human-made disasters are characterised by emotional language expressing anxiety and anger, and by a greater presence of words related to family and friends (Ferron and Massa, 2012).

3.4   Health-Related Content on Wikipedia

Health professionals or laypersons can contribute to Wikipedia's medical and health content. The English-language Wikipedia is a prominent source of online health information as compared to other providers such as MedlinePlus and NHS Direct (Laurent and Vickers, 2009). The Wikipedia's medical content is considered one of the most viewed medical resource globally and was viewed more than 4.88 billion times in 2013 (Heilman and West, 2015). Wikipedia is widely considered a prominent health information source not only by the general public but also by health professionals (Okoli et al., 2014). In fact, 70% of junior physicians use Wikipedia more frequently than all the other websites excluding Google, while nearly 50% to 70% of the practicing physicians use it as an information source to provide medical care (Heilman et al., 2011). Furthermore, Wikipedia is highly used by pharmacists in Australia during work to access general health- and medicine-related information as well as to explore unfamiliar topics (Benetoli et al., 2016).

Taking into consideration that there are concerns around trusting Wikipedia as a health information source (Shafee et al., 2017), various studies have compared the quality of the health-related information on Wikipedia with that of centrally controlled information sources. For example, Candelario et al. (2017) found that Wikipedia medication pages include incomplete and inaccurate information compared with the corresponding United States' medical product guides. Similar studies demonstrated that drug information on Wikipedia is less complete and has more errors than a traditionally edited database such as Medscape Drug Reference (Clauson et al., 2008). Moreover, articles related to complementary and alternative medicines are underdeveloped and require further edits to improve their quality (Koo, 2014). Other studies advised to not rely on Wikipedia as the only source for drug information or dietary supplements as much of the information on Wikipedia is found to be incomplete (Kupferberg and Protus, 2011; Phillips et al., 2014). Researchers have also warned that Wikipedia entries on cardiovascular diseases lack accuracy (Azer et al., 2015). Moreover, Leithner et al. (2010) demonstrated that the quality of osteosarcoma-related information found in the English Wikipedia is inferior to the patient information provided by the US National Cancer Institute (NCI) website. In contrast, other researchers have found that the quality of information on

depression and schizophrenia on Wikipedia is generally as good as, or better than, that on centrally controlled websites such as Encyclopaedia Britannica (Reavley et al., 2012). Along the same lines, research has confirmed that Wikipedia is of sufficient quality to be used as a reference by nursing students (Haigh, 2011). Other studies have concluded that Wikipedia entries in the areas of health, nutrition, medicine, and complementary and alternative medicine are mostly of high quality, but significant errors and omissions are fairly common (Temple and Fraser, 2014). Despite the disagreement over the quality of information on Wikipedia, researchers have suggested that the Wikipedia content related to certain medical conditions can be difficult to understand by patients and their families (Watad et al., 2017; Biggs et al., 2018).

The debate about whether we should trust Wikipedia as a health-information source has not deterred researchers from investigating the patterns and dynamics of health information seeking on Wikipedia. Research has shown that people rely on Wikipedia to search for information about drugs as the highest number of hits was found for pages related to opiates, benzodiazepines, antibiotics, and antidepressants (Law et al., 2011). Dzogang et al. (2016) looked at seasonal fluctuations of the information seeking patterns on Wikipedia and reported that access to articles related to mental health disorders peaks during the winter months. Correspondingly, researchers agreed with prior studies that the search for health information increases at the beginning of the workweek. Researchers have used Wikipedia to monitor the global search volumes for rare conditions such as Systemic Lupus Erythematosus (Sciascia and Radin, 2017). Further, Brigo et al. (2015b) and Brigo et al. (2015a) analysed the number of daily views to Wikipedia articles related to epilepsy, seizures, and other neurological disorders. The studies revealed that the search volume for some of the Wikipedia articles related to a certain disease increased after a news announcement about a celebrity getting diagnosed with the disease. These findings are consistent with those of another study that disclosed that access to articles on rheumatoid arthritis increased after a celebrity's death from rheumatoid arthritis (Mahroum et al., 2018). These studies collectively imply that a celebrity's disease can significantly affect the public health behaviours and lead to increased public interest in the disease's diagnosis or prevention. Besides getting triggered by news about celebrities, Wikipedia search volumes might mirror the patients' fears and worries about particular disorders (Brigo and Erro, 2016).

3.5   Gap in Knowledge

The literature reviewed in this chapter suggests that the role of the Web during crises and emergencies cannot be underestimated. Online forums and social media sites are used during crises by victims and emergency teams to disseminate information and situational updates, collect resources, engage with the audience, and provide emotional

support. Various social media tools, mobile applications, and crowdsourcing applications have been developed to engage with the audience and disseminate information that sometimes could be untrue or unverified. A closer examination of the content on different online forums and social media sites revealed that people experience and express different forms of uncertainty while trying to make sense and interpret information during crises. These forms of uncertainty can be escalated as a result of misinformation and a lack of knowledge that hampers the crisis conditions. Official entities' crisis response and communication can also escalate certain forms of uncertainty. In the context of Wikipedia, page views have been used to measure and assess information seeking behaviours for different articles related to crises. Moreover, the editing activities and the network structure have been evaluated over time to examine the Wikipedia articles related to crises.

On the basis of this literature review, the identification and the filling of knowledge gaps have been recognised and summarised in Figure 3.1. First, there has been limited research on the process of Wikipedia articles' creation during crises. In particular, work that measures and reveals Wikipedia editors' responses to topics related to crises is absent from the literature. This is important to further understand how the element of swift attention that is associated with crisis conditions is conveyed and manifested in Wikipedia articles. This understanding will contribute to the field of Web Science by unveiling different characteristics of how a socio-technical phenomenon such as Wikipedia functions under the conditions of crises in particular and under the effect of the elements of swift attention in general.
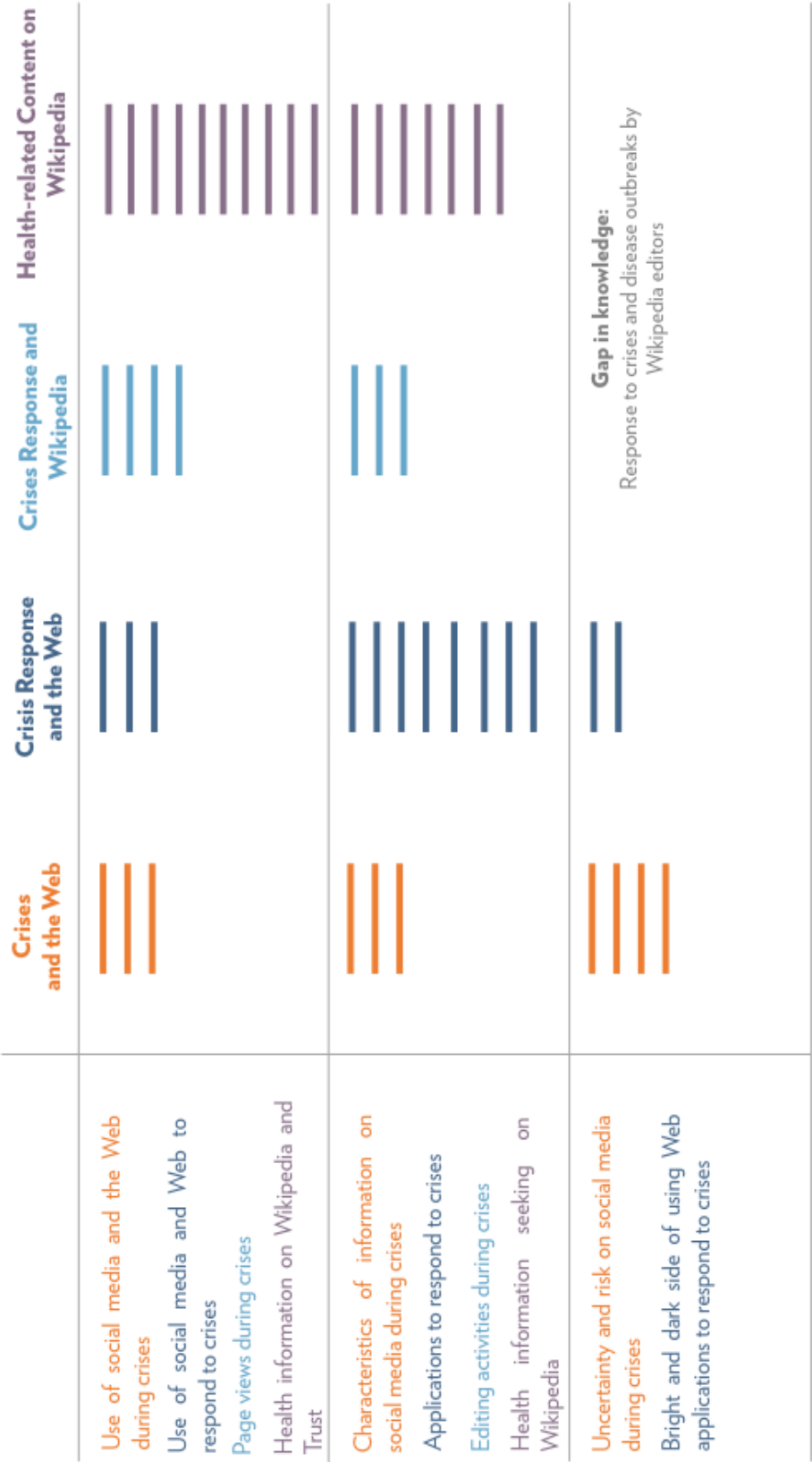
FIGURE 3.1: Emerging Themes: A Review of the Existing Literature and Knowledge Gaps Columns Represent Themes, while Rows Represent Sub-Themes. Cells Display the Number of References Included in Literature Review

3.6   Research Goals and Research Questions

The English Wikipedia includes different articles related to crises and breaking news. Wikipedia's openness, popularity, and up-to-date archival records enable investigating the process of participation and contribution to Wikipedia pages. This contribution can come in many forms and quantities. These forms of contribution include adding new text or posting to talk pages. Prior work has examined dynamics of viewing content on Wikipedia, but has not focused on dynamics of contribution to Wikipedia during crises. Also, previous work has studied specific forms of contributions to Wikipedia articles such as editing activities exclusively, but has not also studied posts to talk pages. Moreover, previous work has looked at the contribution to limited number of Wikipedia articles or topics, but has not considered looking at the contribution to articles related to various crises in comparison to articles that are not related to crises. Our study expands the current understanding of the effects of crises on the contribution to Wikipedia articles. Crises are situations that require immediate attention and interaction between editors in order to create new articles and update constantly changing and uncertain content. Given that previous research disclosed that Wikipedia collaborations around breaking news exhibit unique patterns of activity and engagement within emergent topic spaces (Twyman et al., 2017), this research aims to understand and reveal these patterns by looking at different forms of contribution to Wikipedia articles covering various crises. These forms of contribution include editing activities as well as posts to talk pages. This research also seeks to look at the differences between the contribution to articles related to crises and the contribution to articles that are not related to crises.

This research's goal was to study Wikipedia editors' response to crises in two parts. The first part involved finding patterns of editing behaviours during crises to identify and compare the response to related articles. The second part involved distinguishing among forms and strategies to manage the uncertainty during crises in order to expose the response to the uncertainty. This research focused on the new disease outbreak crises on Wikipedia. This research was also extended to compare and contrast articles related to new disease outbreak crises with articles related to chronic diseases on Wikipedia. This was done to further understand and distinguish the findings related to new disease outbreak crises. Based on the background literature review and the research goals, the overall research question was formulated as follows:

- What effect do crises have on Wikipedia editing?

Three sub-questions were formulated to gather insights about the first part of the research goal:

> RQ1: What are the characteristics of Wikipedia editorial activities during new disease outbreak crises?

RQ2: What are the differences and similarities between editorial activities on Wikipedia articles related to new disease outbreak crises and editorial activities on Wikipedia articles related to chronic diseases?

RQ3: What are the main factors that influence Wikipedia editorial activities during new disease outbreak crises?

Two additional sub-questions were formulated to gather insights about the second part of the research goal:

RQ4: What are the main types of uncertainty that emerge while editing Wikipedia articles related to new disease outbreak crises and Wikipedia articles related to chronic diseases?

RQ5: What are the key Wikipedia editors' strategies for managing uncertainty on articles related to new disease outbreak crises and articles related to chronic diseases?

These questions will help in filling the gap in knowledge by illuminating the process of Wikipedia articles' creation during crises and helping to understand the impact of the elements of swift attention on the contribution to Wikipedia. These questions will be answered using a mixed methods approach that will be described later in the methodology section. The findings will be used to build a framework that helps to model the uncertainty on Wikipedia during crises. Altogether, this thesis makes a conceptual and methodological contribution that will be reported in the discussion section of this thesis.

3.7   Summary

This chapter focused on reviewing the literature on crises and crisis responses on the Web and on Wikipedia. As the focus of this thesis was new disease outbreaks, this chapter also covered literature on health-related content on Wikipedia. A number of issues were raised, emphasising that there has been limited research on the patterns and characteristics of Wikipedia articles' creation during crises. Therefore, the primary research question was formulated to investigate this area followed by five sub-questions. The following chapter moves on to discuss the methodological decisions that were taken to find the answers to these questions.

Chapter 4

Methodology

'An experiment is a question which science poses to Nature, and a measurement is the recording of Nature's answer.'

— Max Planck

The previous chapter introduced this thesis's research questions. Correspondingly, this chapter presents the methodology that was selected to explore these research questions. Firstly, the chapter provides an overview of the case study design that leads to a further discussion on combining case study design with thematic analysis. Considering that the Web Science interdisciplinary research upholds combining qualitative and quantitative methods, the chapter then moves on to presenting the mixed methods research design. In order to provide a rational for selecting a mixed methods research design, this chapter also presents a discussion on the chosen philosophical approach. Thus, the benefits and challenges of using a mixed methods research design are jointly examined. Finally, the research questions are revisited to justify the use of the mixed methods approach.

4.1 Case Study Design

One can define research methodology as the philosophy and framework related to the entire research process (Ahmed et al., 2016). Choosing a particular research methodology is an important step to systematically solve the research problem (Kothari, 2004) and to execute a plan that describes the collection, measurement, and analysis of data (Kothari, 2004). Decisions made about the research design are often influenced by the researchers' capacities, convictions, beliefs, and interests (Ahmed et al., 2016). In addition, decisions made about research design need to address the research questions with the consideration of time, cost, and the resources available (O'Leary, 2004). It is common to select a research design that is based on one or more research purposes such as exploration, description, diagnosis, and experimentation (Kothari, 2004). The

research design also accounts for the type of selected research method, i.e. whether it is quantitative or qualitative research. The research designs for quantitative research are mostly experimental research, ex post facto research, surveys (longitudinal and cross-sectional), and case study. The research design for qualitative research are commonly surveys, action research, exploratory research, grounded theory, ethnography, and case study (Håkansson, 2013). However, it is not unusual to combine one or more types of research design or select a research design that can be applied to both quantitative and qualitative research, such as the case study design (Walliman, 2011).

The case study research design provides an explanatory or descriptive approach for studying elements through a comprehensive analysis of a situation or case. The case study design includes detailed examinations of an individual, group, episode, event, or any other unit of social life (O'Leary, 2004). Following a case study research design involves data collection from a variety of sources such as observations, interviews, audiovisual material, documents, and reports (Creswell, 2014). Investigation also includes a detailed analysis of a bounded system (a case) or multiple bounded systems (cases) over time (Creswell, 2007). In a single or *instrumental* case study, the researcher focuses on a research problem and then, selects a case to provide insights into the problem. Alternatively, in multiple or a *collective* case study, the researcher jointly studies a number of cases to provide insights into the problem (Denzin and Lincoln, 2003). Overall, case study design is often selected to cover the contextual condition of the phenomena under study, particularly when the boundaries between the phenomena and the context are not clearly evident (Yin, 2003). Moreover, case study design is preferred when studying contemporary events and when relevant behaviours cannot be controlled or manipulated (Yin, 2014). On the basis of these factors, case study design was selected as an approach to study the response to crises on Wikipedia. Various contextual conditions related to the editing patterns and the response to the uncertainty on Wikipedia during new disease outbreaks were scrutinised in this thesis.

There are a number of limitations of using a case study as a research design. Case studies have been criticised for their lack of rigour as researchers tend to have a biased interpretation of the data (Zainal, 2007). Moreover, findings generated by following the case study design can be difficult to generalise beyond a small number of cases and subjects (Gerring, 2007). Nevertheless, generalisability could be achieved by examining multiple cases and multiple events (Noor, 2008). Certain approaches of case study design allow researchers to combine quantitative and qualitative data sources and gain a holistic view about a phenomenon, population, or general condition.

4.2 Case Study Research and Thematic Analysis

One of the shortcomings of using the case study research design is the lack of generalisation or prediction (Dooley, 2002). Nevertheless, the lack of generalisation does not deter researchers from applying case study research to build thematic frameworks. In particular, a theme is usually defined as a "coherent integration of the disparate pieces of data that constitute the findings" (Vaismoradi et al., 2013). Themes can offer insights into patterns or meanings across a data set. These patterns of meaning are captured to explore certain research topics and answer the research questions. After identifying and reviewing each theme, a theoretical framework can be built to visually present main themes, sub-themes, and interconnections between themes and sub-themes. The theoretical framework can be developed using thematic analysis (Vaismoradi et al., 2013).

Thematic analysis is a process of "identifying, analyzing, and reporting patterns (themes) within data" (Castleberry and Nolen, 2018). Thematic analysis is a qualitative method that can be widely used to analyse large qualitative data sets and answer across a range of epistemologies and research questions (Nowell et al., 2017). Thematic analysis provides detailed analysis of meanings within a particular context (Vaismoradi et al., 2013). Thematic analysis can be applied as a top-down or bottom-up approach. The top-down approach or theoretical thematic analysis is driven by existing concepts or theories, while the bottom-up or inductive analysis that is driven by the data itself (Maguire and Delahunt, 2017). Thematic analysis is an iterative process that involves moving back and forward between phases (Nowell et al., 2017).

Different prominent analysis methods such as discourse analysis and content analysis can be applied to analyse textual information. Discourse analysis mainly focuses on the analysis of language including face-to-face talk, non-verbal interaction, images, symbols and documents to study social life, culture, and linguistics (E Shaw and Bailey, 2009). However, content analysis focuses on analysing textual information by systematically coding and categorising data to reveal trends and patterns (Brough, 2018). Although that content analysis is very similar to thematic analysis as both aim to code textual data, the two methods have important differences. Content analysis is classified as both a quantitative and qualitative method, while thematic analysis is classified as a purely qualitative method (Brough, 2018). Accordingly, thematic analysis is applied in this research because it has several key advantages. Thematic analysis offers researchers a flexible and complex analysis approach that can be modified to the needs of different studies (Nowell et al., 2017). Thematic analysis is useful in examining similarities and differences in perspectives within data and generating unanticipated insights (Nowell et al., 2017). Even though that thematic analysis is quick to learn in comparison to other qualitative research methods, there is a lack of literature on thematic analysis in

relative to literature on grounded theory, ethnography, and content analysis. In addition, the flexibility of thematic analysis can lead to inconsistency and lack of coherence when developing themes derived from the research data (Nowell et al., 2017).

4.3   Mixed Methods Approach

There are three types of research design: qualitative, quantitative, and mixed methods. The qualitative approach aims to explore a topic of interest by collecting data through case studies, ethnographic work, and interviews (Harwell, 2011). Qualitative research attempts to explore people's thoughts and perspectives and interpret phenomena in terms of the meanings that people bring to them (Harwell, 2011). In contrast, the quantitative approach aims to explore a topic of interest by collecting data through tests, questionnaires, and surveys (Harwell, 2011). Quantitative research assumes that there is only one 'truth' and attempts to apply statistical methods to maximise the objectivity, replicability, and generalisability of the findings (Hesse-Biber, 2010; Harwell, 2011). In contrast, the mixed methods approach aims to combine quantitative and qualitative research techniques and methods. The core assumption is that combining methods provides a more complete understanding of the research problem that would not be possible by relying on either approach alone (Johnson et al., 2007; Creswell, 2014). This is mainly because the mixed methods approach merges the strengths and offsets the weaknesses of the quantitative and the qualitative approaches. Furthermore, the mixed methods approach provides comprehensive evidence and helps to answer questions that quantitative or qualitative methods alone could not answer. The mixed methods approach is important today because of the complexity of the problems that need to be solved and the practical need to gather multiple forms of data for a diverse audience (Creswell, 2007).

As indicated in Table 4.1, there are four types of mixed methods design, namely embedded, explanatory, exploratory, and triangulation. Embedded research design involves collecting data sequentially so that one dataset acts to support the other dataset. However, explanatory and exploratory research designs involve collecting data in two stages. Researchers who follow explanatory design collect quantitative data and then collect qualitative data to explain the quantitative results. However, researchers who follow exploratory research design collect data in the reverse order where quantitative data are collected to explain the qualitative results. Finally, triangulation research design involves collecting both quantitative and qualitative data at the same time to merge the results and generate greater reliability (Ahmed et al., 2016).

The triangulation design was therefore adopted as explained above. The triangulation

Design Type	Implementation	Priority	Stage of Integration
Embedded	Concurrent collection of quantitative and qualitative data	Quantitative or qualitative	Analysis phase
Explanatory	Quantitative followed by qualitative	Usually quantitative; can be qualitative or equal	Interpretation phase
Exploratory	Qualitative followed by quantitative	Usually qualitative; can be quantitative or equal	Interpretation phase
Triangulation	Concurrent collection of quantitative and qualitative data	Preferably equal; can be quantitative or qualitative	Interpretation phase or analysis phase

TABLE 4.1: Types of Mixed Methods Design (Creswell and Clark, 2007)

design is defined as a "combination of methodologies in the study of the same phenomenon" (Jick, 1979). Triangulation allows researchers to use qualitative and quantitative methods and corroborate different results related to the same phenomenon. Triangulation is used when researchers want to directly compare and contrast quantitative results with qualitative results or to validate or expand quantitative results using qualitative data (Creswell and Plano Clark, 2011). This research applies a specific type of triangulation mixed methods design: the convergence model. The convergence model is applied to collect and analyse quantitative and qualitative data on the same phenomenon separately and then converge (compare and contrast) the results. The purpose of using this model is to collect diverse sources of data to answer the research questions and reach a comprehensive conclusion. The triangulation method is illustrated in Figure 4.1.
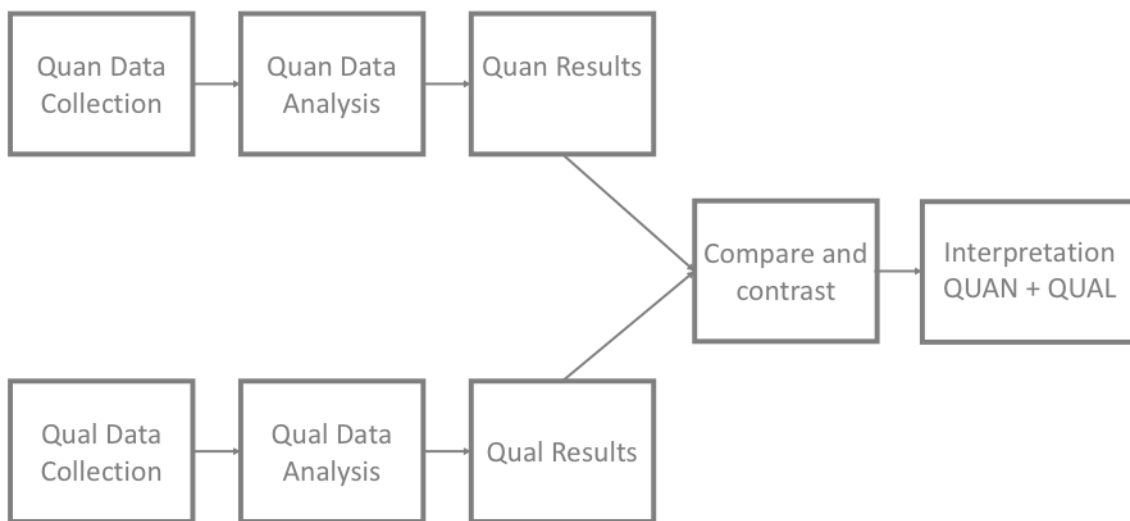
FIGURE 4.1: Convergence Model of Triangulation Design
(Creswell and Plano Clark, 2011)

4.4   Mixed Methods Approach Paradigm

The choice of a research methodology is often influenced by the researchers' convictions, beliefs, and interests. These beliefs can be described as paradigms or worldviews that hold the philosophical assumptions to guide the research actions (Creswell, 2007). There are a wide variety of paradigms that can be associated with different research methods. For example, quantitative research is often associated with the positivism paradigm. The positivism paradigm argues that reality is independent of the researcher and must be investigated using a rigorous process of empirical inquiry. Therefore, empirical inquiry should be based upon observations and quantitative measures to test a hypothesis or analyse the casual relationships between variables (Gray, 2014; Golafshani, 2003). In general, positivism and quantitative research regard the world as made up of observable facts and work together to collect data to uncover laws and general patterns (Punch, 2014).

In contrast, qualitative research is often associated with interpretivism. The interpretivism paradigm believes that reality does not exist in the external world, but it is created by the subject's views and interactions with the world (Gray, 2014). Interpretive studies concentrate on the meanings that subjects bring to certain situations (Punch, 2014). Hence, interpretive studies allow multiple or contradictory views to the world phenomena to exist (Gray, 2014). These views are usually captured by qualitative approaches to data gathering and analysis such as case studies and ethnography (Thanh and Thanh, 2015). Qualitative approaches offer interpretivists the opportunity to fully understand the context as well as to understand the complex and changing reality (Thanh and Thanh, 2015).

After selecting the mixed methods approach for this research, we sought a philosophical paradigm that accepts the insights provided by both qualitative and quantitative methods (Johnson and Onwuegbuzie, 2004). Along these lines, we argued that the pragmatism paradigm is regarded as the best way to guide our research. Pragmatism is defined as a research philosophy that aims to build bridges between conflicting philosophies and shed light on how research approaches can be mixed fruitfully (Johnson and Onwuegbuzie, 2004). The pragmatism paradigm "recognizes the existence and importance of the natural or physical world as well as the emergent social and psychological world that includes language, culture, human institutions, and subjective thoughts" (Johnson and Onwuegbuzie, 2004). Holding this worldview allows researchers to focus on the research problems and the outcomes of the research, rather than on the antecedent conditions and the methods. Therefore, the pragmatism worldview gives researchers the freedom to choose the methods, techniques, and procedures of research that best meet their needs and purposes. The pragmatism paradigm allows selecting either the positivism and the interpretivism position or combining both these positions within the scope of a single research. Thus, the pragmatism paradigm argues that reality should be captured by using both quantitative and qualitative sources of data collection (Creswell, 2007). Accordingly, the pragmatism worldview is considered a suitable paradigm for conducting mixed methods research.

4.5 Benefits and Challenges of Mixed Methods Approach

Various disciplines are traditionally linked with specific research paradigms (i.e. engineering and computer science are often associated with positivism). However, interdisciplinary research often requires researchers to look beyond a specific research paradigm and use a variety of methods (Doyle et al., 2009). This research takes advantage of the mixed methods approach to answer a broad range of our interdisciplinary research's questions and develop a comprehensive and clear understanding of the research problem. This is done by demanding corroboration between the quantitative and the qualitative data and explaining certain phenomena using qualitative and quantitative findings. For example, findings from surveys (quantitative) can be further explained by conducting interviews (qualitative) (Doyle et al., 2009). Moreover, the mixed methods approach provides strong and accurate inferences by using the strengths of one method to neutralise and overcome the weaknesses of the other method (Creswell et al., 2003). These characteristics of the mixed methods approach can be applied to achieve greater validity and stronger evidence than those obtained using either a qualitative or a quantitative approach (Ahmed et al., 2016; Doyle et al., 2009).

Although the mixed methods approach has much to offer, researchers should be careful when using this approach in their research. Mixed methods research has been criticised

for holding incompatible ontological and epistemological assumptions that cannot be merged into a single study (Creswell, 2011). This criticism arises as a result of blending different data collection and research methods that are usually linked to certain paradigms (Creswell, 2011). Others have criticised the pragmatism philosophy for supporting mixed methods research. This is done by arguing that making methodological choices solely on the basis of pragmatism or on the basis of what works does not answer the question "practical for whom and to what end?" (Doyle et al., 2009). In addition, applying a mixed methods approach can be expensive and time consuming in terms of collecting and analysing quantitative and qualitative data (Ahmed et al., 2016). Despite these drawbacks, one cannot deny that social phenomena are complex and that different types of methods are needed to best understand these complexities (Creswell et al., 2003). This approach is particularly required to understand the complexities that exist in socio-technical phenomena under the field of Web Science.

4.6   Research Questions and Mixed Methods Approach

This thesis used a mixed methods research design to understand the Wikipedia editors' response to crises and new disease outbreaks. The quantitative research approach was applied to reveal and compare the patterns of editing behaviours during crises as underlined by Q1, Q2, and Q3. In particular, Q1 aimed to describe the patterns of editing behaviours on Wikipedia during new disease outbreak crises. Therefore, a quantitative approach was applied because it allows a large-scale data analysis on the number of edits of Wikipedia articles. Moreover, Q2 aimed to compare the patterns of editing behaviours on Wikipedia articles related to new disease outbreaks with the patterns of editing behaviours on Wikipedia articles related to chronic diseases. Accordingly, a quantitative approach was applied, because it facilitates comparisons and helps in "distinguishing characteristics, elemental properties, and empirical properties and tends to measure how much and how often" (Amaratunga et al., 2002). Further, Q3 aimed to look at factors that could influence the patterns of editing behaviours on Wikipedia during new disease outbreak crises. As such, a quantitative approach was applied because it allows conducting a statistical analysis to establish a correlation between the editing activities on Wikipedia articles and the other factors or outcomes.

However, a qualitative research approach was applied to reveal the forms and strategies to manage the uncertainty during crises as highlighted in Q4 and Q5. In particular, Q4 aimed to disclose the different types of uncertainty expressed by Wikipedia editors during new disease outbreak crises. Thus, a qualitative approach was applied because it is well-suited for understanding the meanings that people place on events or processes (Amaratunga et al., 2002). In particular, the qualitative approach emphasises on "people's lived experience" (Amaratunga et al., 2002), which makes it ideal for studying

Wikipedia editors' experience with uncertainty during new disease outbreak crises. In addition, Q5 aimed to reveal strategies that Wikipedia editors apply to manage the uncertainty during new disease outbreak crises. Accordingly, the qualitative approach was applied because it provides "rich descriptions that are vivid and nested in a real life context" (Amaratunga et al., 2002). Taking into consideration that this thesis focuses on describing strategies to manage the uncertainty in the context of crises and new disease outbreaks, qualitative approaches made it possible to gain a condensed and broad description of the contextual elements. Altogether, this thesis required collecting data over an extended period of time in order to trace how editors express and manage the uncertainty during new disease outbreaks crises. Correspondingly, qualitative approaches were powerful in examining processes or procedures, because the qualitative data were predominantly collected over an extended period of time (Amaratunga et al., 2002).

The quantitative and qualitative findings were triangulated and discussed to gain insights about the Wikipedia editors' response to crises. Therefore, this thesis applied a mixed methodology research design to develop a framework that depicted multiple forms of uncertainty and strategies to manage the uncertainty during crises on Wikipedia. A mixed methods approach combines quantitative and qualitative data where "qualitative data can support explicitly the meaning of quantitative research" (Amaratunga et al., 2002). As such, the quantitative approach helped to assess and distinguish the editing patterns in the sample articles, while the qualitative approach helped to develop themes to explain and understand the context of a response to the uncertainty in the sample Wikipedia articles. Combining both quantitative and qualitative approaches helped to merge the strengths of both these approaches and provided a rich understanding of the Wikipedia editors' response to new disease outbreak crises.

4.7   Summary

This chapter provides a complete description of the methodological approach used in this thesis. Case study research design was applied to explore the Wikipedia editors' response to crises. As introduced in Chapter 3, this response can be expressed in two different forms. A mixed methods approach was used to examine the research questions under these forms. First, a quantitative research method was used to find the editing patterns during crises. Second, a qualitative research method was used to distinguish the forms and strategies to manage the uncertainty. Subsequently, the qualitative and quantitative findings were triangulated in order to build a framework that aimed to describe the uncertainty on Wikipedia during crises.

Chapter 5

The Case Study: Disease Outbreaks and Chronic Diseases Wikipedia Articles

'The important thing in science is not so much to obtain new facts as to discover new ways of thinking about them.'

—Sir William Lawrence Bragg

Building on the case study design and the mixed methods approach described in the previous chapter, this chapter explains the selected case study, data, and analytical methods used in this thesis. These methods help to find the patterns of editing behaviours during crises so as to identify and compare the responses to the related articles (RQ1, RQ2, and RQ3). Furthermore, these methods help to distinguish forms and strategies to manage the uncertainty during crises so as to expose the response to the uncertainty (RQ4 and RQ5). This chapter begins by discussing the criteria used to select the case, which is then followed by a review of the qualitative and quantitative methods chosen to analyse data. The chapter also illustrates the process of integrating qualitative and quantitative findings. The chapter ends by illustrating ways to ensure the quality of the applied methods.

5.1 Defining the Case Study

The previous chapter discussed that the case study research design is applied to study the Wikipedia editors' responses to crises. Accordingly, multiple or collective cases have been selected to provide a comprehensive view and generalise the research findings. These cases are Wikipedia articles related to new disease outbreaks. 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"(World Health Organization, 2019c). 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, 2019c). Health events related to new diseases are usually announced by the World Health Organization (WHO)[1]. The World Health Organization reports the number of cases, mortality rates, and warnings of global risk.

In general, new disease outbreaks represent uncertain situations because of the limited information and scientific knowledge about these diseases. 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 dynamics. 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 might cause stigma and public panic, mainly spread by the media (Garcia Serpa Osorio-de Castro et al., 2017).

In contrast, control cases are selected to compare and distinguish the findings related to new disease outbreaks. These control cases are Wikipedia articles related to chronic diseases. Chronic diseases or non-communicable diseases (NCDs) are diseases that are not passed from person to person. They are of a long duration and generally progress slowly (World Health Organization, 2019a). Active research to tackle problems related to chronic diseases is conducted around the world. Nevertheless, different forms of uncertainty related to the diagnosis, prognosis, and treatment of chronic diseases still exist.

5.2   Sample Wikipedia Articles

Certain Wikipedia articles on new disease outbreaks and chronic diseases have been selected. In particular, Wikipedia articles related to five new disease outbreaks that occurred in the past ten years were selected. These articles were the ones on the Zika virus, Ebola, swine influenza, avian influenza (H7N9), and Middle East respiratory syndrome (MERS). Hence, in all, 52 articles covering these five new disease outbreaks were retrieved from Wikipedia. In contrast, Wikipedia articles related to five chronic diseases were selected. These articles were the ones on diabetes, obesity, cancer, asthma, and cardiovascular diseases. These diseases were selected because they are the leading causes of death worldwide. Cardiovascular diseases account for most of the NCD deaths, or 17.9 million people annually, followed by cancers (9.0 million), respiratory diseases (3.9 million), and diabetes (1.6 million) (World Health Organization, 2019b). In all, 153 articles related to diabetes, obesity, cancer, asthma, and cardiovascular diseases were retrieved from Wikipedia. These Wikipedia articles are selected in order to capture all the articles

---

[1]https://www.who.int/about

covering the selected diseases. Out of the 205 Wikipedia articles that were included in the sample, 10 articles covering each disease were identified as the main articles and are listed in Table 5.1. These articles were chosen as the main articles because: 1) they had the highest number of revisions; 2) these articles show up on the top of Google's search engine results as demonstrated in Figures 5.1 to 5.10. Therefore, these articles are likely to get more attention than other articles and considered the most frequently edited articles.

Article's Title	Article's Page Views
Zika virus	8,074,442
Ebola virus disease	1,374,374
Middle East respiratory syndrome-related coronavirus	12,559
Influenza A virus subtype H7N9	25,625
Swine influenza	660,66
Diabetes mellitus	494,815
Obesity	1,597,954
Cancer	3,127,971
Asthma	1,614,093
Cardiovascular disease	980,456

TABLE 5.1: Main Wikipedia Articles and Articles's Page Views as retrieved from https://tools.wmflabs.org/pageviews/ from 01/07/2015 to 31/12/2016
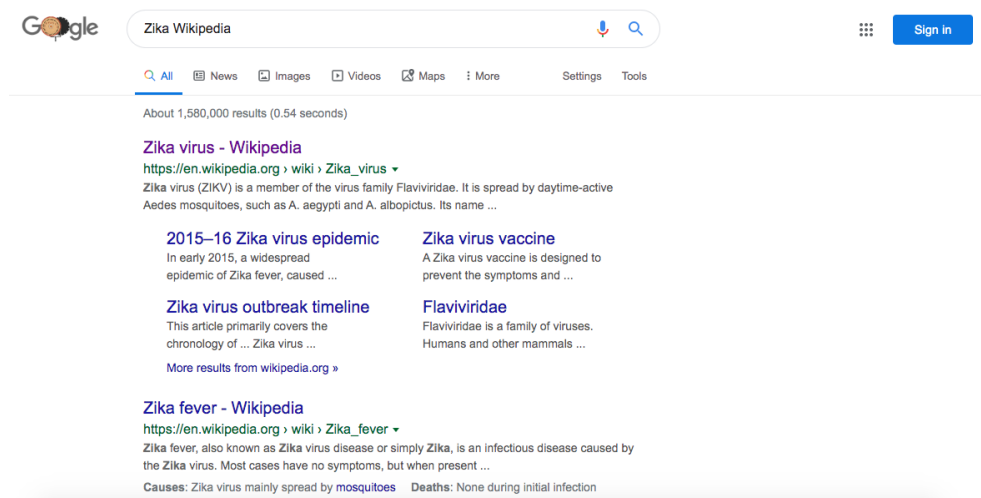
FIGURE 5.1: Zika Search Results
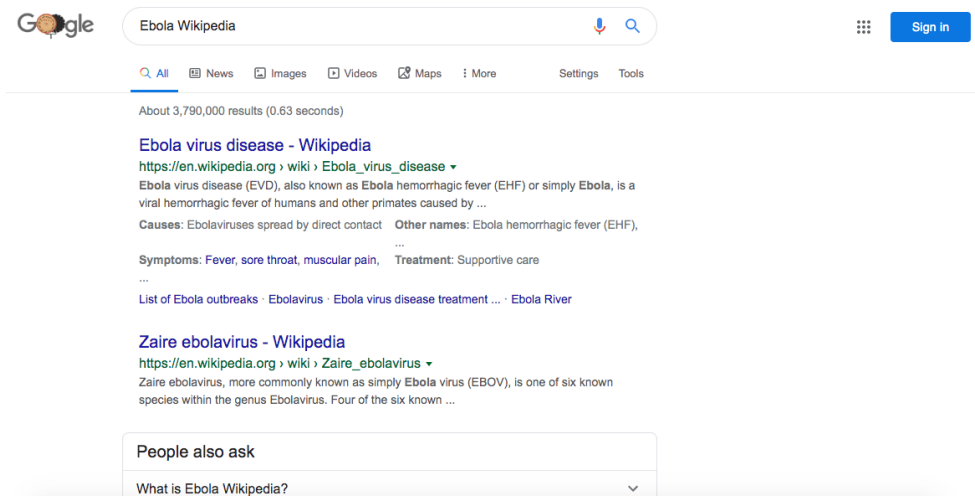
FIGURE 5.2: Ebola Search Results
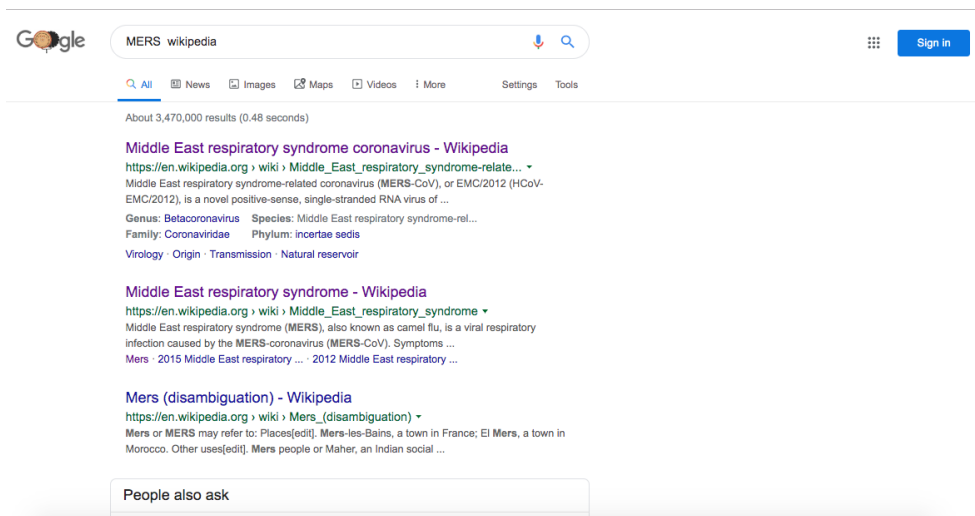
FIGURE 5.3: MERS Search Results
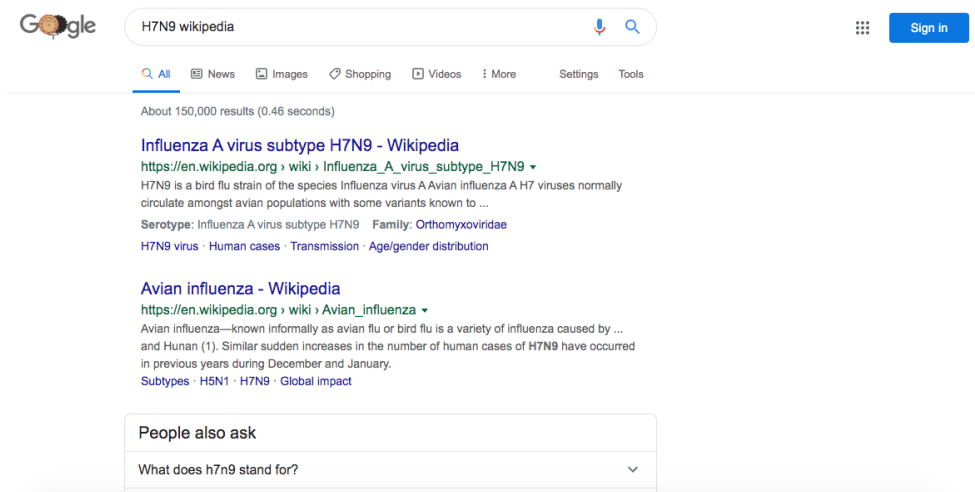
FIGURE 5.4: H7N9 Search Results
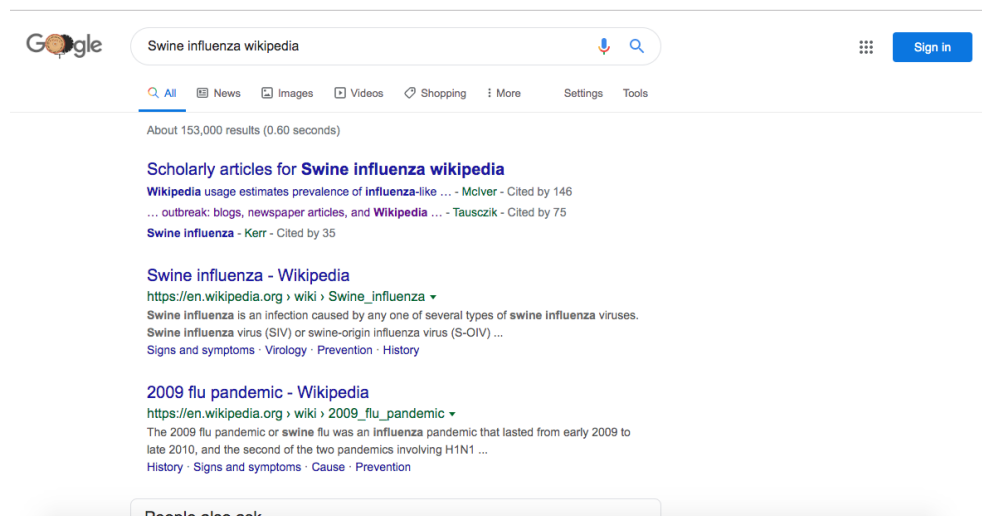
FIGURE 5.5: Swine Influenza Search Results
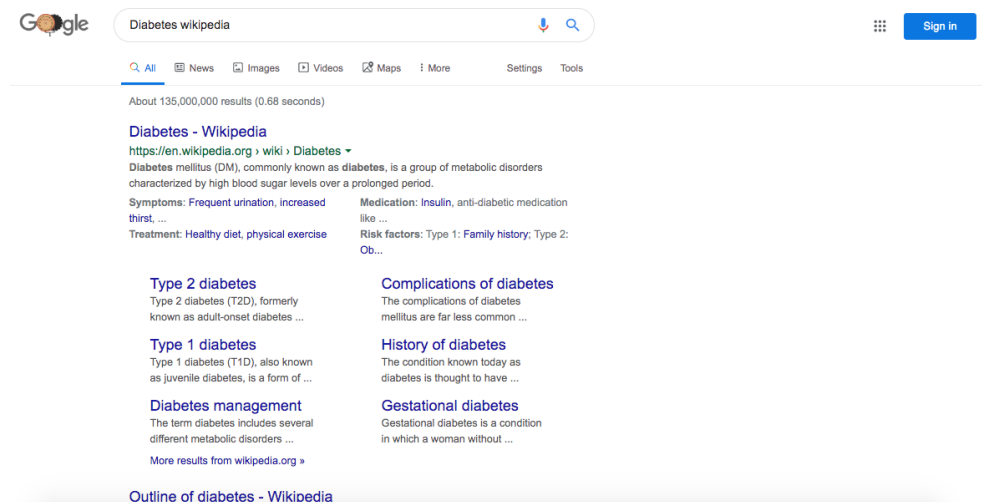
FIGURE 5.6: Diabetes Search Results
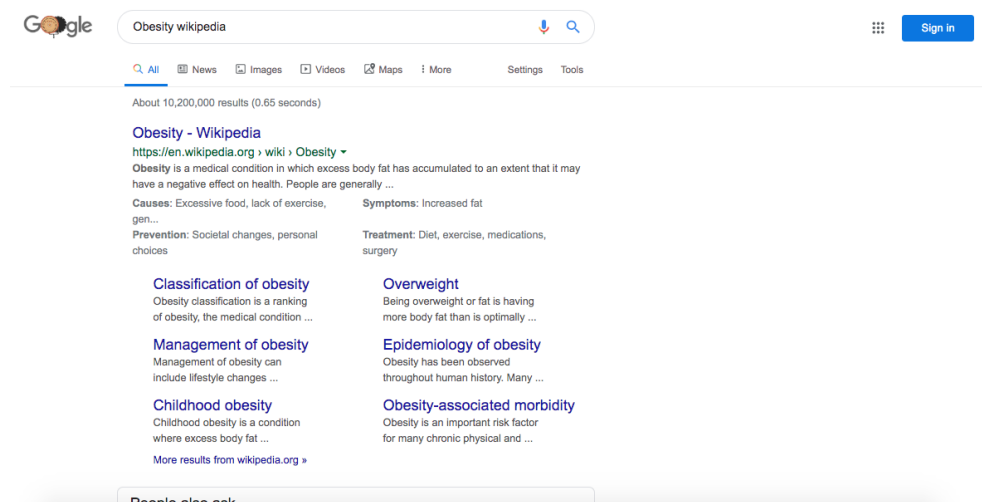
FIGURE 5.7: Obesity Search Results
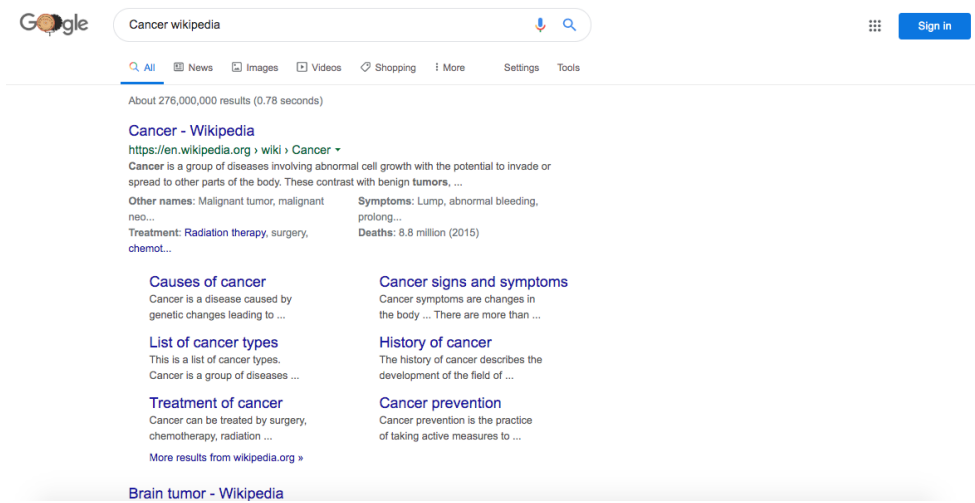
FIGURE 5.8: Cancer Search Results
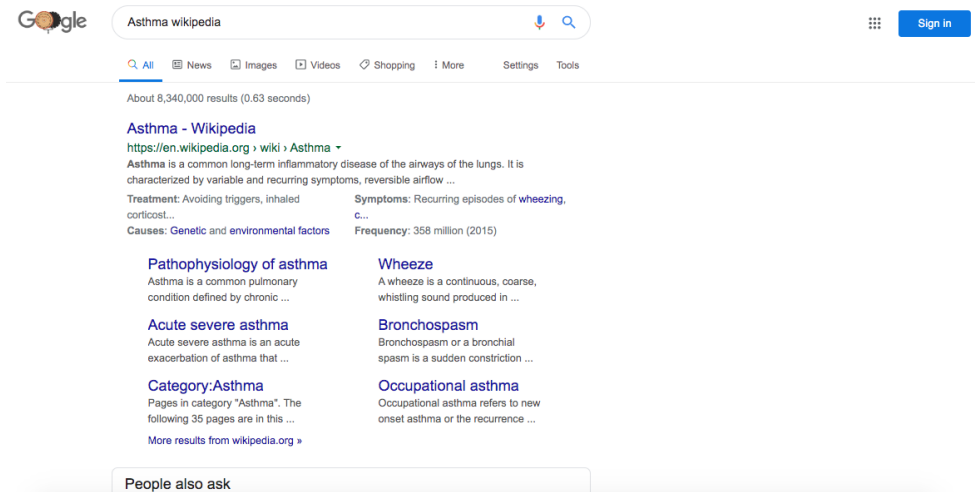
FIGURE 5.9: Asthma Search Results
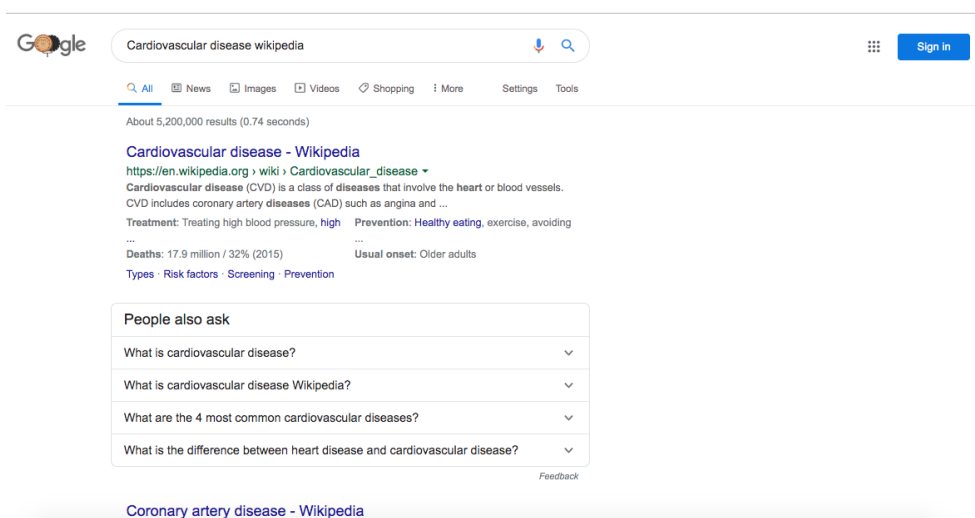
FIGURE 5.10: Cardiovascular Disease Search Results

5.3 Quantitative and Qualitative Data Sources

Quantitative data sources were obtained from the Wikipedia revision history. The data were harvested from the Wikipedia API[2], covering all the revisions of the 205 selected Wikipedia articles from 2001 to 2016. Each revision included an ID, username, timestamp, and comment on the change made. Usernames included bots and anonymous users, marked as 'anon' in the entry. The following is an example of a revision metadata retrieved from the Wikipedia API:

revid="724203877" parentid="724051463" user="Seungbae007" timestamp="2016-06-07T19:46:15Z" comment="uploading an image"

Qualitative data sources were obtained from Wikipedia talk pages' archives. The talk pages' archives were selected because they include discussions between editors in the form of posts and replies to these posts. The content of these discussions can unfold how editors expressed and managed uncertainty during new disease outbreaks. Thus, posts on the talk pages of the 10 main articles were collected, covering all the content from 2001 to 2018. This produced a total of 938 talk pages (261 pages related to articles on new disease outbreaks and 677 pages related to articles on chronic diseases). These pages include a total of 1686 posts (477 posts on pages related to new disease outbreaks and 1209 posts on pages related to chronic diseases). The following is an example of a Wikipedia post from the Zika article:

> *I am not convinced this is needed in the first sentence. It is not the most important thing about the condition. Was at the end of the lead which IMO is better. Doc James (talk · contribs · email) 04:10, 21 July 2016 (UTC)*

5.4 Research Questions and Analytical Analysis

As discussed in the previous section, the quantitative research methodology was used to explore Q1, Q2, and Q3. Q1 aims to describe patterns of editing activities on Wikipedia during crises. These editing activities consist of two data sets: the number of revisions per day and the number of unique editors per day. Therefore, the selected methods should enable to reveal the given norm or average of these two data sets as well as data or observations that deviate drastically from the given norm or average of these two data sets. Consistently, outlier detection methods were selected because these methods can assess every observation and detect observations that tend to be inconsistent with the majority of the observations in a dataset (Tang et al., 2007; Hodge and Austin, 2004). Outliers detection methods have been applied to study two different phenomena:

---

[2]https://en.wikipedia.org/w/api.php

anomaly and burstiness. Anomaly is the percentage of outliers that are detected in *either* the data set of the number of revisions per day *or* the data set of number of unique editors per day. However, burstiness is the percentage of outliers that are *simultaneously* detected in *both* the data set of the number of revisions per day *and* the number of unique editors per day. For example, certain days could have outliers detected in the data set of the number of revisions per day, but not in the data set of number of unique editors per day (anomaly). Also, certain days could have outliers detected in both the data set of the number of revisions per day and the data set of number of unique editors per day (burstiness). Outlier detection methods help to discover exceptional behaviours in certain phenomena and accordingly was applied to discover exceptional editing activities in Wikipedia during crises.

Furthermore, the head/tail breaks classification scheme was selected as a method to compare the patterns of editing behaviours on Wikipedia articles (Q2 and Q3). As will be demonstrated in Chapter 6, the patterns of editing activities on Wikipedia had a heavy-tail distribution. Therefore, the head/tail breaks classification scheme was selected because it assists in finding different classes or hierarchical levels for data with a heavy-tail distribution (Jiang, 2015b). The measurement of the correlation between different variables was used as a statistical technique to explore the relationship between the editing activities on Wikipedia articles and the other factors or outcomes (Q3).

In contrast, the qualitative research approach was used to disclose the forms and strategies to manage the uncertainty during crises (Q4 and Q5). As highlighted in the previous section, there is no existing theoretical grounding or thematic framework that looks at the response to the uncertainty on Wikipedia during crises. As a result, the thematic analysis was used to develop a thematic framework that discloses forms and strategies to manage the uncertainty on Wikipedia during crises. This is because the thematic analysis is appropriate for building frameworks that can explain a phenomenon in a specific context (Cho and Lee, 2014). Therefore, a conventional qualitative analysis was applied by coding and deriving themes directly and inductively from the data (Zhang and Wildemuth, 2009). Quantitative and Qualitative methods will be discussed in detail throughout this chapter.

5.5   Quantitative Analysis

A quantitative analysis was performed to detect the outliers in the dataset of the Wikipedia revision history. Different methods of outlier detection are introduced in this section. These methods are Z-score, Tukey's method, local outlier factor (LOF), seasonal hybrid ESD (S-H-ESD), breakout detection, and autoregressive integrated moving average (ARIMA). Ground-truth data were labeled to evaluate the accuracy of each

outlier detection method. Testing different outlier detection methods is important in order to: 1) detect and compare the percentage of anomalies found in the data set of the number of revisions per day or in the data set of number of unique editors per day; 2) select the method that differentiate between editing activities on Wikipedia articles related to new disease outbreak crises and editing activities on Wikipedia articles related to chronic diseases; 3) select the most accurate method to measure burstiness. The concept of burstiness is introduced in relation to the existing literature in this section. As it will be demonstrated in Chapter 6, the distribution of both the number of revisions and the number of unique editors is skewed to the right. Therefore, the head/tail break classification scheme was used to classify the level of burstiness in Wikipedia articles because this scheme has been originally developed to classify data that is positively or negatively skewed. Taking this into consideration, this section also introduces the concept of the head/tail break in relation to the current literature.

5.5.1 Outlier Detection

An outlier is typically defined as "an observation that deviates so much from other observations as to arouse suspicion that it was generated by a different mechanism" (Ben-Gal, 2005). Outliers could occur as a result of the measurement errors or as indicators of important occurrences or events. Taking into consideration that outlier detection and analysis is an important and challenging data mining task, a variety of outlier detection techniques have been developed to find patterns in data that do not conform to the expected behaviour (Singh and Upadhyaya). This research explores the detection of outliers in the dataset of the articles' revision history using six different outlier detection techniques, namely Z-score, Tukey's method, LOF, S-H-ESD, breakout detection, and ARIMA.

5.5.1.1 Z-score

The z-score is a statistical transformation that specifies the number of standard deviations that a value is above or below the mean (Lavrakas, 2008). A z-score can be calculated using the following formula:

$$z = (X - \mu)/\sigma \tag{5.1}$$

where z is the z-score, X is the value of the element, μ is the population mean, and α is the standard deviation. A negative z-score means that an observation is below the mean, while a positive one means that it is above the mean. Z-scores that are further away from 0 indicate that the observations are also further away from the mean. Usually, outliers are identified when an observation falls in more than three standard deviations from the mean (Shiffler, 1988).

5.5.1.2 Tukey's Method

Tukey's method is used to detect outliers for both symmetric and skewed datasets. Turkey's method identifies outliers by calculating the data quartiles as points that are either:

$$\text{Below } Q1 - 1.5\text{IQR or above } Q3 + 1.5\text{IQR} \qquad (5.2)$$

The first quartile *Q1* is the value $\geq 1/4$ of the data, the second quartile *Q2* or the median is the value $\geq 1/2$ of the data, and the third quartile *Q3* is the value $\geq 3/4$ of the data. The interquartile range *IQR* is Q3 − Q1 (CMU Statistics, 2013).

5.5.1.3 Local Outlier Factor (LOF)

The LOF algorithm is usually applied to detect outliers irrespective of the data distribution. The main idea is that the density around an outlier significantly differs from the density around its neighbours (Auskalnis et al., 2018). As such, the LOF method scores each data point by computing the ratio of the average densities of the point's neighbors to the density of the point itself. The estimated density of a point is the number of neighbors divided by the sum of distances to the point's neighbors (Turi Machine Learning Platform User Guide, 2019):

$$(\text{p}) = \frac{k}{\sum_{x \in N(p)} d(p,x)} \qquad (5.3)$$

where *N(p)* is the set of neighbors of point *p*, *k* is the number of points in this set, and *d(p,x)* is the distance between points *p* and *x* (Turi Machine Learning Platform User Guide, 2019). Based on this equation, the LOF score is calculated as follows:

$$\text{LOF(p)} = \frac{\frac{1}{k} \sum_{x \in N(p)} \hat{f}(x)}{\hat{f}(p)} \qquad (5.4)$$

5.5.1.4 Seasonal Hybrid (ESD S-H-ESD)

This method works by combining seasonal decomposition and the generalized ESD test, for anomaly detection (Hochenbaum et al., 2017). Seasonal decomposition is used to determine the seasonal component of a given time series, while the generalized ESD test is used to identify the anomaly data by their distance from the mean value (Wu et al., 2018). The S-H-ESD algorithm had initially been applied on Twitter data (Hochenbaum et al., 2017) and was implemented later as an open-source R package to detect anomalies in the presence of a seasonality or an underlying trend (Twitter, 2019a).

5.5.1.5 Breakout Detection

This method works without requiring the data to follow any specific distribution. Breakout detection intends to detect changes in the time series by detecting the mean divergence and the change in the data distribution. Mostly, the algorithm detects the divergence by applying a mean clustering technique. The breakout detection algorithm was also released by Twitter as an open-source R package (Twitter, 2019b).

5.5.1.6 ARIMA

The ARIMA model was developed by Box and Jenkins in 1970 (Box and Jenkins, 1990). This model can detect the outliers in the time series data. The model combines the autoregressive model (AR) and the moving average model (MA) with non-seasonal differences added to the model. AR uses the dependent relationship between an observation and some of the lagged observations. The integrated model (I) uses the differencing of observations (e.g. subtracting an observation from an observation in the previous time step) to make the time series stationary. While MA uses the dependency between an observation and a residual error from the MA applied to the lagged observations (Brownlee, 2017; Hyndman and Athanasopoulos, 2018). The model can be expressed as follows:

$$\hat{y}_t = \mu + \Phi_1 Y_{t-1} + ... + \Phi_p Y_{t-p} - \Theta_1 \epsilon_{t-1} - \Theta_q \epsilon_{t-q} \tag{5.5}$$

The ARIMA model can have different parameters ARIMA(p,d,q) that are defined as follows:

- **p**: Number of lagged observations included in the model, also called the lag order.

- **d**: Number of times that the raw observations are differenced, also called the degree of differencing.

- **q**: Size of the moving average window, also called the order of moving average (Brownlee, 2017; Hyndman and Athanasopoulos, 2018).

Accordingly, this research tested the ARIMA model using different parameters as summarised as below (Duke University, 2019):

1. ARIMA(1,0,0) = first-order autoregressive model:

$$\hat{y}_t = \mu + \Phi_1 Y_{t-1} \tag{5.6}$$

2. ARIMA(0,1,0) = random walk model:

$$\hat{y}_t = \mu + Y_{t-1} \tag{5.7}$$

3. ARIMA(1,1,0) = differenced first-order autoregressive model:

$$\hat{y}_t = \mu + Y_{t-1} + \Phi_1(Y_{t-1} - Y_{t-2}) \tag{5.8}$$

4. ARIMA(0,1,1) without constant = simple exponential smoothing:

$$\hat{y}_t = Y_{t-1} + \Theta_1\epsilon_{t-1} \tag{5.9}$$

5. ARIMA(0,2,1) or (0,2,2) without constant = linear exponential smoothing:

$$\hat{y}_t = 2Y_{t-1} - Y_{t-2} - \Theta_1\epsilon_{t-1} - \Theta_2\epsilon_{t-2} \tag{5.10}$$

6. ARIMA(1,1,2) without constant = damped-trend linear exponential smoothing

$$\hat{y}_t = Y_{t-1} + \Phi_1(Y_{t-1} - Y_{t-2}) - \Theta_1\epsilon_{t-1} - \Theta_1\epsilon_{t-1} \tag{5.11}$$

5.5.2   Ground-Truth Data

The Wikipedia revision history of the Zika virus article was chosen randomly as the ground-truth data. Two annotators manually labelled the outliers as observed in the number of revisions per day. The inter-annotator agreement was measured using Fleiss' Kappa score to find the reliability and the consistency between the annotations. Fleiss' Kappa is presented in equations 5.12, 5.13, and 5.14 and can be interpreted as the extent to which the observed amount of agreement among the raters exceeded the expected amount of agreement if all the raters made their ratings completely randomly (Schaer et al., 2010). Therefore, Fleiss' Kappa score of the Zika virus article's annotations was 0.83 which was a perfect agreement score.

$$K = \frac{\overline{P} - \overline{P}_e}{1 - \overline{P}_e} \tag{5.12}$$

where

$$\overline{P} = \frac{1}{Nn(n-1)} \sum_{i=1}^{N} \sum_{j=1}^{K} n_{ij}(n_{ij} - 1) \tag{5.13}$$

and

$$\overline{P}_e = \sum_{j=1}^{K} (\frac{1}{Nn} \sum_{i=1}^{N} n_{ij})^2 \tag{5.14}$$

5.5.3   Accuracy

With the ground-truth data, the accuracy rate was used as a metric to evaluate the different outlier detection models. Simply, the accuracy rate is the fraction of correct

predictions and can be calculated as follows:

$$\text{Accuracy} = \frac{\text{Number of correct detections}}{\text{Total number of detections}} \qquad (5.15)$$

5.5.4 Burstiness

The most accurate model was adopted to measure the burstiness in the data set of the number of revisions per day or in the data set of number of unique editors per day. Burstiness is generally defined as a sudden and brief increase in *something*, or a short appearance of something (Dictionary, 2018). In particular, burstiness is short timeframes of intense activity or frequency of an event followed by long times of no or reduced activity (Goh and Barabasi, 2006; Akbarpour and Jackson, 2018). Burstiness can be observed in a broad range of phenomena, such as in email patterns, Web browsing, and network traffic, and even in natural events such as earthquakes and gene expressions (Goh and Barabasi, 2006). In addition, burstiness can be observed in text streams when there is a rapid increase in the frequency of a single term or phrase (Kalogeratos et al., 2016). Therefore, methodological approaches that measure burstiness have been applied in different areas of research such as to identify emerging scientific research areas (Guo et al., 2011) and to rank topics from different online news streams (Wang et al., 2008).

Web Science research has benefited from the development of methodological approaches that detect and measure burstiness. Primarily, search engines index and query the underlying collections by applying frequency-based measures. These measures record the frequency of a term in each document, typically normalised by a global frequency measure (Lappas et al., 2009). Further research looked at the potential of identifying temporal burstiness by being aware of frequency-based measures in specific moments of time. In particular, a temporal burst can be captured by finding a) a time interval that spots the occurrence of burstiness; and b) a score of the burst's strength, which is typically recorded by measuring the extent of deviation from the usual frequency (Lappas et al., 2013a). Accordingly, different methods and approaches have been recommended to detect temporal burstiness. For example, moving average can be applied to time series data to detect burstiness. This is by considering values that are above specific cut-off points of the standard deviation above the mean (Piotrkowicz et al., 2017; Vlachos et al., 2004). Further, parameter-free methods have been advanced to detect temporal burstiness without relying on parameter tuning and estimation. The parameter-free approach was utilised to identify the bursty features in a text stream and accordingly, group strongly interrelated bursty features as events (Fung et al., 2005). Furthermore, the parameter-free approach was coupled with the discrepancy theory for the purpose of identifying burstiness in the time intervals for any given term (Lappas et al., 2009). The discrepancy theory describes the deviation of a situation from the 'expected' behavioural

baseline and is applied in several fields including machine learning and computer graphics (Lappas et al., 2009). Moreover, research has considered finding more information about the data distribution such as skewness - the asymmetry of the data distribution - and kurtosis - the degree of peakedness of the distribution - to adjust the threshold values for real-time burst detection (Khaing and New, 2017).

Research has been extended to detect the burstiness linked to trends and real-world events such as natural disasters and health alerts. For example, Guzman and Poblete (2013) introduced an approach for online bursty keyword detection to be potentially used to identify the emerging topics on Twitter. This approach relies on analysing the window slicing and the window relevance variation rate. In particular, the approach normalises keyword frequencies by relevance and compares them in adjacent time windows. This approach also investigates the different characteristics related to stopword identification such as stopword signals should be noisier (i.e. have a higher standard deviation) than the signals of regular keywords (Guzman and Poblete, 2013). Similar work has reported a method to discover meaningful events from online content. This includes detecting an event by capturing the basic frequency of a set of words that are associated with this event from the posts on a microblogging site. Moreover, the approach combines both Latent Dirichlet allocation[3] and Jenssen-Shannon divergence[4] to measure the similarity of the detected events (Lu et al., 2014). Other work attempted to further enhance the event detection methods by developing additional measures that can facilitate the identification of bursty terms and combine them into a single measure. These measures include skewness (bursty terms have a thicker tail distribution), consistency (bursty terms' occurrences across different channels), periodicity (bursty terms are less likely to be periodic), and variation (reducing the possibility of identifying bursty terms if they are bursty only in a specific channel) (Lee et al., 2012). In addition, other research has explored the use of optimisation models[5] to correlate multiple activities that drive event-related bursts on social media (Zhao et al., 2012). Other work has focused on mining both temporal and spatial burstiness patterns from streams such as online and geo-aware microblogging platforms (Lappas et al., 2013b). Moreover, Kotozaki (2015) proposed a novel algorithm based on the kernel density estimation[6] that detects location-based bursts. The algorithm allows to extract the peak periods and locality of trending topics and events. Finally, Schubert et al. (2014) applied outlier detection methods for locating the emerging topics early, long before they become hot topics.

---

[3]is a probabilistic generative model that can be used to estimate the properties of multinomial observations and explain the similarities between some observations (Heinrich, 2005).

[4]is a method of measuring the similarity between two probability distributions (Heinrich, 2005).

[5]These models rely on the idea that activity streams related with one query might be dependent, i.e., the states of multiple activity streams on the same timestamp tend to be the same. Even though that various activity streams are different in the scale of frequencies, they tend to share similar trend patterns (Zhao et al., 2012).

[6]is a method that estimates the overall distribution from the location information dataset (Kotozaki, 2015).

5.5.5   Head/Tail Break

Given the fact that the distribution of both the number of revisions and the number of unique editors follow heavy tailed distributions as it will be demonstrated in Chapter 6, this thesis proposes to classify the burstiness on Wikipedia by using the head/tail breaks classification scheme. The head/tail breaks classification scheme was recently developed to classify data with heavy-tailed distributions such as power law, log-normal, and exponential distributions (Jiang, 2015b; Chen, 2014; Li et al., 2016; Jiang and Ma, 2015). The head/tail method relies on measuring the arithmetic mean to split the data into the head (the small percentage of values above the mean) and the tail (the large percentage of values below the mean) (Ma et al., 2017). The breaking process proceeds recursively for the head part, until the head values in the last iteration are no longer a minority and no longer have a heavy-tailed distribution (Jiang, 2015a; Jiang and Sui, 2014). Thus, this method guarantees that the resulting number of classes, as well as the class intervals are naturally and automatically derived on the basis of the inherent hierarchy of the data (Jiang, 2015a; Schiller, 2016). The number of classes or hierarchical levels is called the ht-index (h) and can be measured as follows (Jiang and Okabe, 2014):

$$h = m + 1 \tag{5.16}$$

where m records the maximum number of times that the head/tail breaks method performed iteratively on the head part of a dataset (Gao et al., 2017). Recording the value of m is important to measure different dimensions such as the complexity and the heterogeneity of the data in hand. In other words, higher values of the ht-index indicate more complexity in the hierarchical structure (Jiang, 2015b), as well as more heterogeneity in the data. Accordingly, it is possible not only to use the ht-index to analyse the data complexity but also to detect heavy-tailed distributions (Ma et al., 2017).

The head/tail breaks method has several advantages as compared to other classification schemes such as the Natural Breaks classification (or optimisation). First, the number of classes is objectively and naturally determined, rather than subjectively imposed as the Natural Breaks classification (Jiang and Sui, 2014). In particular, the head/tail breaks method relies on an automatically defined mean that acts recursively as a cut-off to differentiate classes. In contrast, the Natural Breaks classification defines classes on the basis of a pre-defined mean and variance (Jiang, 2015a). Second, the head/tail breaks are effective in capturing the hierarchical structure for the data with a heavy-tailed distribution. This is in contrast to the Natural Breaks classification scheme, which often fails to capture the scaling structure of the data with a heavy-tailed distribution (Jiang and Sui, 2014). Third, the head/tail breaks give a more correct expression of the data's topographical distribution, so it provides a better representation of the number of classes than the Natural Breaks classification scheme. This was confirmed by a study that tested and compared the head/tail breaks with natural breaks in order to classify

the elevations above (heights) and below (depths) the sea level in the US (Lin, 2013). However, the study also acknowledged that the Natural Breaks scheme performed better than the head/tail breaks when the numbers of classes is large (i.e. above 16 classes) (Lin, 2013). Overall, a majority of studies agree that the head/tail breaks is considered a simple, transparent, easily reproducible, and well-suited method for data with a heavy-tailed distribution (Dunham-Snary et al., 2014; Li et al., 2016; Jiang, 2015b).

The head/tail breaks method has been mostly applied to capture the hierarchy of geographical features. This is because the method works well in determining class intervals for a cartographic representation, generalising geographic features, and visualising the city structure and evolution (Gao et al., 2016). The head/tail breaks method has also been explored as a solution to the problem of spatial heterogeneity (Gao et al., 2016), as well as a classification method to divide and map the levels of disaster risk (Fariza et al., 2015). Taking into consideration that the head/tail breaks method can be applied to classify big data, the method has been utilised as a classification method for online data (Ma et al., 2017). For example, Ma et al. (2015) has utilised the head/tail breaks as a classification method for the number of users, elements, and contributions on OpenStreetMap. The use of the head/tail breaks approach was justified because the numbers of users, elements, and contributions exhibit a heavy-tailed distribution (Ma et al., 2015). Moreover, the head/tail breaks approach has been used to examine the spatial distribution of tweets and densities in order to better understand the city structure and dynamics (Jiang et al., 2016). Equivalently, after collecting the location data from Brightkite and Gowalla (location-based social media sites), Ma et al. (2017) used the head/tail breaks approach to classify check-in locations and patterns in order to understand the social and geographical aspects of human activities.

5.6   Qualitative Analysis

A qualitative analysis was performed to uncover the different sources of uncertainty and strategies to manage an uncertainty as expressed by the editors in the Wikipedia talk pages. The posts on the talk pages of the articles related to new disease outbreaks and chronic diseases have been explored. The results have been used to build a thematic framework of the uncertainty on Wikipedia during new disease outbreak crises.

5.6.1   Thematic Analysis

As proposed in Chapter 4, the thematic analysis was applied to analyse the qualitative data. The following subsections summarises the six-phase process for thematic analysis:

5.6.1.1 Phase 1: Becoming Familiar with the Data

The first step in any qualitative analysis is reading, and re-reading posts on the talk pages of the articles related to new disease outbreaks and chronic diseases (Maguire and Delahunt, 2017). This phase involves storing the data in a well-organized archive in NVivo and documenting initial thoughts about the data (Nowell et al., 2017) as shown in Figure 5.11.
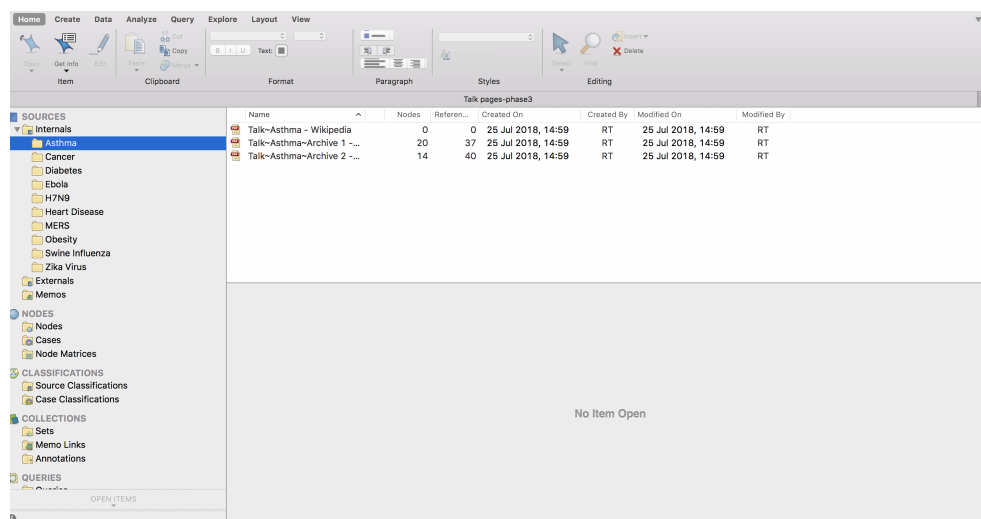
FIGURE 5.11: Archives of Talk Pages in NVivo

5.6.1.2 Phase 2: Generating Initial Codes

The analysis of qualitative results was initiated by coding the data. Coding is "the process of generating ideas and concepts from raw data" (Given, 2008), whereas codes refer to "concepts and their identification through explicit criteria" (Given, 2008). Systematically, codes can reveal interesting meanings and features of the entire data set.

Table 5.2 shows the research's codes that have been generated inductively after going through all the posts and assessing each post to identify relevant aspect to the research questions (Nowell et al., 2017). NVivo facilitates the coding process by allowing the creation of nodes to capture important sections and attach index labels to text. Accordingly, all the codes are created and stored in NVivo with the references to the coded text as shown in Figure 5.12.

Code	Description
Scientific Uncertainty	Uncertainty is expressed on talk pages when scientific information is reported in publications as uncertain or if relevant publications are either absent or poor
References Uncertainty	Uncertainty is expressed on talk pages about information or claims mentioned in particular references or publications. This is expressed when trying to clarify the claims within the references, checking whether certain references are valid, or simply questioning claims coming from some references.
Conflicting References Uncertainty	Uncertainty is expressed on talk pages when coming across conflicting and incompatible claims mentioned in multiple publications as well as in literature in general
Lack of References Uncertainty	Uncertainty is expressed on talk pages when coming across claims that have no references
Manage Uncertainty	Strategies used to manage uncertainty include relying on authoritative sources, reporting the uncertainty in the article, ignoring the uncertainty, consulting experts for advice, and setting up a mailing list about ongoing research
Controversy	Controversies are identified as any discussion on Wikipedia talk pages that includes disagreement between more than two editors (Appendix D)
Controversy Type 1	Disagreement between editors about claims related to scientific and technical knowledge. This includes disagreement about the interpretation of scientific information
Controversy Type 2	Disagreement between editors about whether certain references can be trusted
Request for Information	Posts that include questions or requests for clarifications and information
Add Information	Posts that include requests to consider adding additional information to articles

Continued on next page

Continued from previous page

Code	Description
Check for Vandalism	Posts that include requests to check vandalism on Wikipedia. Vandalism is the addition, removal, or modification of text or other material that is either humorous, nonsensical, a hoax, offensive, or degrading in nature[1]
Simplify Articles	Posts that include requests to simplify the language used to write Wikipedia articles so as to make the articles easier to read and understand
Recentism	Posts that include warnings against inflated or imbalanced focus on recent events[2]
Conflict of Interest	Posts that include warnings against adding content on Wikipedia about yourself, family, friends, clients, employers, or financial information[4]
Original Research	Posts that include warnings against posting materials with no reliable or published sources exist[3]
Synthesis of Publications	Posts that include warnings against combining materials from multiple sources to reach or imply a conclusion that is not explicitly stated by any of the sources[5]

Table 5.2: Initial Codes

5.6.1.3 Phase 3: Searching for Themes

This phase involves sorting and collating all the initially coded data into relevant themes. A theme is an abstract entity that identifies patterns of meaning on Wikipedia talk pages posts. As such, each theme "captures and unifies the nature or basis of the experience into a meaningful whole" (Nowell et al., 2017). Therefore, five different themes have been created based on the coded data. These themes are: *controversy*, *requests*, *strategies to*

---

[1] https://en.wikipedia.org/wiki/Vandalism_on_Wikipedia
[2] https://en.wikipedia.org/wiki/Wikipedia:Recentism
[4] https://en.wikipedia.org/wiki/Wikipedia:Conflict_of_interest
[4] https://en.wikipedia.org/wiki/Wikipedia:No_original_research
[5] https://en.wikipedia.org/wiki/Wikipedia:No_original_research#Synthesis_of_published_material

FIGURE 5.12: Nodes in NVivo

manage uncertainties, *uncertainty*, and *warnings*. To organise these themes, NVivo allows nodes (codes) to have more than one dimension (tree branch). Sorting themes as parent nodes in the tree structure Figure (5.12) facilitates identifying common properties and making early comparison. These identified themes can connect data pieces together and capture important aspects to the proposed research questions (Q4 and Q5).

5.6.1.4 Phase 4: Reviewing Themes

This phase involves applying the emerged themes to the entire posts on Wikipedia talk pages to determine if there is enough data under each theme to answer the research questions. The review process includes redefining, splitting, or discarding some of the preliminary themes (Brough, 2018). During this phase, it becomes evident that the *controversy*, *requests*, and *warnings* do not have enough data to support them or provide insights to address the research questions. This leads to eliminating these three themes. Besides, two sub-themes under the uncertainty theme have been merged together. Specifically, the *lack of references uncertainty* sub-theme has been merged with the *scientific uncertainty* sub-theme. Additionally, the strategies to manage uncertainties theme has been broken down to five different sub-themes as to identify and describe each strategy. Table 5.3 describes the revised codes, while Figure 5.13 shows the refined themes and sub-themes on NVivo. At the end of this phase, themes appear to form a more coherent pattern to tell a clearer story about posts on articles related to new disease outbreaks and chronic diseases.

Code	Description
Scientific Uncertainty	Uncertainty is expressed on talk pages when scientific information is reported in publications as uncertain or if relevant publications are either absent or poor
References Uncertainty	Uncertainty is expressed on talk pages about information or claims mentioned in particular references or publications
Conflicting References Uncertainty	Uncertainty is expressed on talk pages when coming across conflicting and incompatible claims mentioned in multiple publications as well as in the literature in general
Relying on Authoritative Sources	Relying on sources that originate from authorities such as the CDC and the WHO to clear doubts and address various forms of uncertainty
Reporting the Uncertainty	Reporting uncertain statements in articles so as to explicitly inform readers about various forms of uncertainty
Ignoring the Uncertainty	Neglecting the uncertain information, and, in effect, editing it until more certain knowledge unfolds
Consulting Experts for Advice	Asking domain experts such as doctors or health professionals to clarify the uncertainty
Setting up a Mailing List	Setting up a mailing list to track the updated research and information to address the ongoing uncertainty

Table 5.3: Revised Themes

5.6.1.5 Phase 5: Defining and Naming Themes

This phase involves conducting analysis to refine each theme and examine how each theme leads to answering the research questions (Brough, 2018). This analysis starts with reading through the data and scrutinizing the themes and sub-themes at least twice (Nowell et al., 2017). Moreover, this analysis includes looking at the interactions and relationships between themes and sub-themes (Maguire and Delahunt, 2017). We found different functions of NVivo particularly useful in conducting the analysis at this phase.

FIGURE 5.13: Revised Themes in NVivo

These functions are: coding queries and matrix coding queries. Coding queries have been used to view all the references coded at each theme and sub-theme. This includes viewing all the references coded under each type of uncertainty and strategy to manage uncertainty. For example, Figure 5.14 shows the results of the coding queries to retrieve all the references coded under uncertainty across all the talk pages in our sample. These references were further used and presented in Table 8.3 and Table 8.4.

However, matrix coding queries have been used to search for intersecting themes and sub-themes (i.e. search for text coded to multiples nodes simultaneously). Matrix coding queries have been used to search for the intersection between themes and sub-themes under uncertainty. For example, matrix coding queries helped to search for references that were simultaneously coded as scientific uncertainty and references uncertainty as is shown in Figure. The results of this query have been summarised in Figure 8.2. In addition, matrix coding queries have been used to search for intersection between themes and sub-themes under strategies to manage uncertainties. For instance, matrix coding queries helped to search for references that were concurrently coded as report the uncertainty in the article and consult experts for advices as is shown in Figure 5.15. The results have been summarised in Figure 8.11. After refining all the main themes, a thematic framework has been developed in Figure 8.12 to illustrate the relationship between themes and sub-themes.

5.6.1.6   Phase 6: Writing-up

The goal of the final phase is to carry out the final analysis and write-up a report that could be a journal article or dissertation (Brough, 2018). The write-up of the analysis should provide a detailed description of the themes that reflects the overall story of the
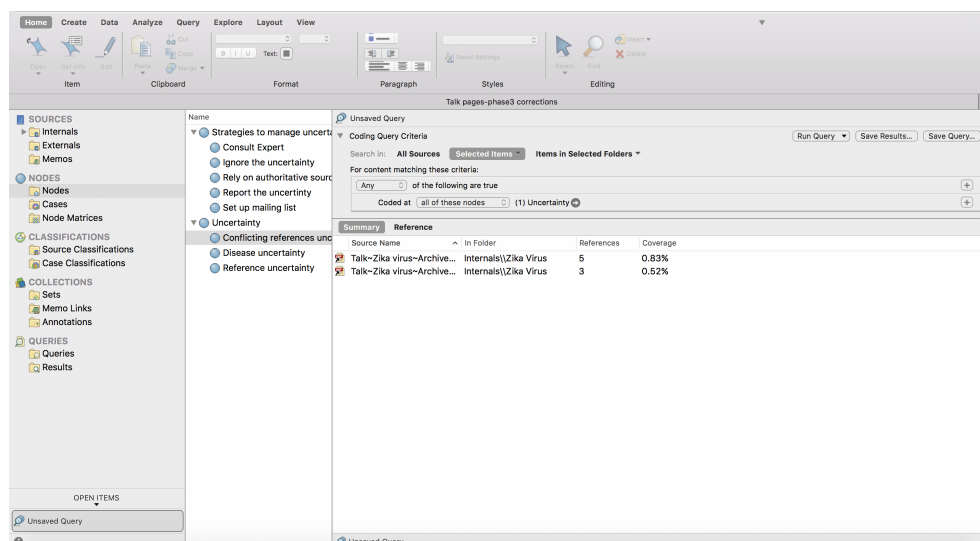
FIGURE 5.14: Results of Coding Query in NVivo
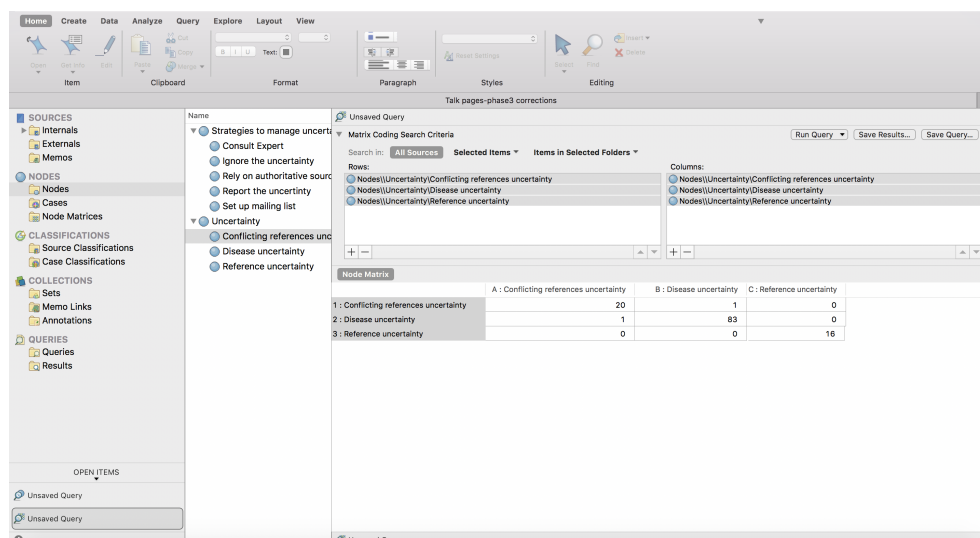
FIGURE 5.15: Results of Matrix Coding in NVivo

data in relationship to the research question. The write-up of the thematic analysis
on Wikipedia talk pages of the articles related to new disease outbreaks and chronic
diseases has been presented in Chapter 8 and Chapter 9. This phase involves embedding
direct codes and references from data as an essential component of the final report
(Nowell et al., 2017). As such, examples of types of uncertainty and strategies to manage
uncertainty have been incorporated in Chapter 8. Furthermore, this phase involves
relating the analysis and interpretation to the literature (Brough, 2018). Consequently,
existing literature on information uncertainty was reviewed and summarised in Appendix
B in order to compare and contrast the developed themes with other work. Besides,
existing literature related to forms and strategies to manage uncertainty have been
reviewed and discussed in Chapter 9. The final thematic framework has been presented
in Figure 9.2 after triangulating the quantitative and qualitative findings and linking

the findings to existing theories.

5.7   Ensuring Rigour and Quality of Methods

Even though the mixed methods approach is usually applied to provide a rich and co-herent understanding of the Web (Tinati et al., 2012), it is also important to magnify the reliability and the validity of the quantitative and qualitative research findings. Reliability refers to the extent to which the results can be consistent over time and can be reproduced using a similar methodology (Golafshani, 2003). The reliability of the results can be ensured through an audit trial such as by describing how the quantitative and qualitative data are collected and analysed (Zohrabi, 2013). Therefore, Python and R scripts were used to retrieve data, keep records, and analyse the quantitative data. In addition, NVivo was used to code and analyse the qualitative data iteratively.

In contrast, validity refers to whether the research truly measures what it was intended or claimed to measure (Golafshani, 2003). Validity is particularly vital in qualitative work where results and interpretations are prone to the researchers' biases and subjectivity (Brink, 1993). Thus, this research benefited from several ways to strengthen the validity of the data evaluation and findings: triangulation and long-term observations (Zohrabi, 2013). Note that the findings from the Wikipedia revision history and talk pages were merged and compared over an extended period of time. In addition, multiple cases related to new disease outbreaks and chronic diseases were selected to increase our confidence in the generalisability of the study.
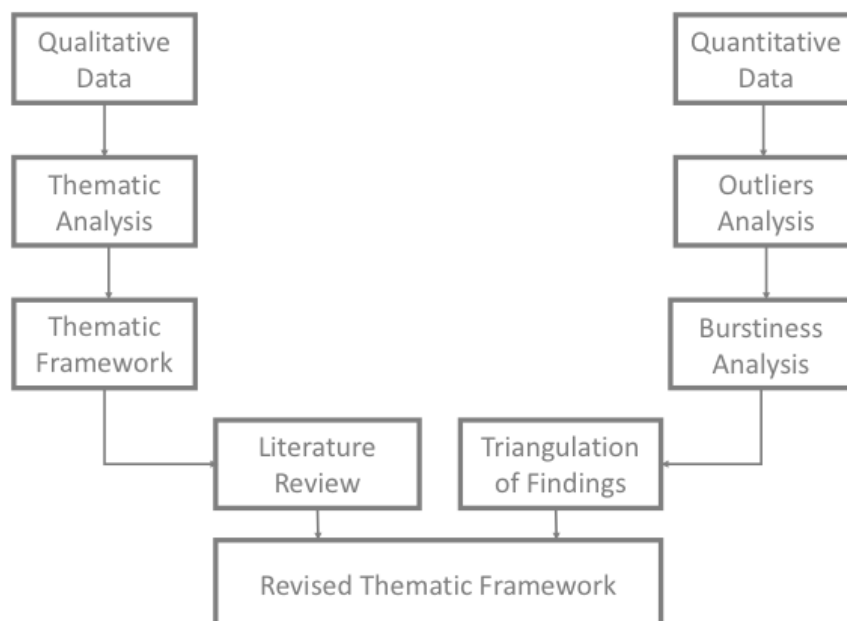
FIGURE 5.16: Framework Building

5.8   Summary

This chapter described the case study used to collect and analyse Wikipedia data. Different quantitative and qualitative methods were introduced to carry out the investigation and to answer the research questions. The next chapter presents a quantitative analysis by exploring the Wikipedia revision history and applying different outlier detection methods.

Chapter 6

Results: Dynamics and Deviations in Wikipedia Editing Activities

'We learn from every natural disaster. Whether it's a fire or a flood, we learn something from it so we can respond to the next one better.'

— Malcolm Turnbull
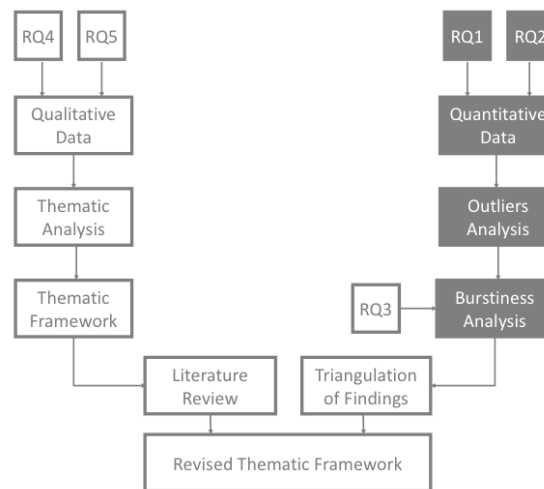
Figure 6.1: Chapter's Placement within Wider Research Design

This chapter aims to describe and compare the patterns of editing behaviours in articles related to new disease outbreaks with editing behaviours in articles related to chronic diseases. This was done by presenting the quantitative findings in response to RQ1 and RQ2. Particularly, the goal of Q1 is to reveal the characteristics of Wikipedia editorial activities during new disease outbreak crises, while the goal of Q2 is to reveal

the differences and similarities between editorial activities on Wikipedia articles related to new disease outbreak crises and editorial activities on Wikipedia articles related to chronic diseases. These questions were answered by studying two different phenomena: anomaly and burstiness. These phenomena were explored by applying the outlier detection methods and the head/tail breaks clustering scheme. This chapter starts by presenting an overview of the Wikipedia edit history and data distribution. First, different outlier detection models were assessed in order to 1) detect and compare the percentage of anomalies found in the data set of the number of revisions per day or in the data set of number of unique editors per day; 2) select the method that differentiate between editing activities on Wikipedia articles related to new disease outbreak crises and editing activities on Wikipedia articles related to chronic diseases; 3) select the most accurate method to measure burstiness in articles related to new disease outbreaks and in articles related to chronic diseases. Second, the head/tail breaks clustering scheme was applied to classify the burstiness in Wikipedia articles into different levels. This chapter concludes by differentiating the results found in articles related to new disease outbreak crises. These results help to understand the dynamics and deviations in editing behaviours during crises.

6.1   Wikipedia Edit History

The previous chapter indicated that quantitative data sources were retrieved from the Wikipedia revision history. An overview of the total number of articles, revisions, and editors is presented in Table 6.1. These quantitative data are further assessed and examined throughout this chapter.

Articles	Number of Articles	Number of Revisions	Number of Editors
Zika Virus	4	4633	1270
Ebola Virus Disease	10	14134	4851
MERS	4	1800	687
H7N9	2	2925	1364
Swine Influenza	32	26303	8572
Diabetes	16	16872	7699
Obesity	6	11738	4729
Cancer	56	53893	22674
Asthma	7	7518	3050
Cardiovascular Disease	68	26897	12968
New Disease Outbreaks	52	49795	16744
Chronic Diseases	153	116918	51120
Total Number of Articles	205	166713	67864

TABLE 6.1: Total Number of Wikipedia Articles, Revisions, and Editors

6.2 Data Distribution

As a fundamental step to describe the patterns of editing behaviours on Wikipedia during new disease outbreak crises (RQ1) and compare the patterns of editing behaviours on Wikipedia articles related to new disease outbreaks with the patterns of editing behaviours on Wikipedia articles related to chronic diseases (RQ2), it is crucial to understand the underlying data distribution. Therefore, to characterise the shape of the distribution of the number of edits and the number of editors per day for each article, the Shapiro-Wilk and Kolmogorov-Smirnov normality tests were conducted. The skewness and kurtosis values were also measured. The results of these tests were used to either accept or reject the null hypothesis that the data is normally distributed.

Table 6.2 reports the results showing that the number of edits and the number of editors per day were well-modeled by a normal distribution. The small p-value (<0.05) suggested that the null hypothesis that the data are normally distributed was rejected. Further, all the distributions had positive kurtosis and skewness values, which indicated that the distributions were very leptokurtic and right skewed. The same tests were conducted on the number of revisions per day and the number of editors per day of the 205 articles in our dataset. The results of the normality tests of all the distributions were statistically significant with a p-value of (<0.05). This led us to conclude that the distributions were leptokurtic and tended to be skewed to the right. The same results were obtained after removing the days that did not receive any new revisions (i.e. the number revisions per day was equal to zero) from the analysis. The positively skewed or skewed right distribution of the data reflects that the right tail of the distribution is longer than the left. The points that are clustered at the right tail of the distribution

imply that Wikipedia articles usually receive a small number of revisions per day as well as attract a small number of editors per day. However, the points that are clustered at the left tail of the distribution imply that Wikipedia articles could receive a higher than usual number of revisions per day or a higher number of editors per day. This implies that editing activities could deviate from the normal editing patterns. The following section looks at detecting these exceptional cases by examining the outliers in articles related to new disease outbreaks and articles related to chronic diseases.

Articles	Shapiro-Wilk Test	Kolmogorov-Smirnov Test	Kurtosis Value	Skewness Value
Zika Virus:				
Revisions	W = 0.5 *	D = 0.5 *	17.3	3.6
Editors	W = 0.5 *	D = 0.5 *	25.2	4.3
Ebola Virus Disease:				
Revisions	W = 0.4 *	D = 0.5 *	122.9	9.0
Editors	W = 0.6 *	D = 0.5 *	50.1	5.3
MERS:				
Revisions	W = 0.5 *	D = 0.5 *	32.7	4.9
Editors	W = 0.5 *	D = 0.5 *	32.4	4.5
H7N9:				
Revisions	W = 0.6 *	D = 0.5 *	11.0	2.7
Editors	W = 0.7 *	D = 0.5 *	12.8	2.7
Swine Influenza:				
Revisions	W = 0.2 *	D = 0.5 *	132.4	10.4
Editors	W = 0.2 *	D = 0.5 *	133.1	10.4
Diabetes:				
Revisions	W = 0.7 *	D = 0.5 *	16.6	2.9
Editors	W = 0.8 *	D = 0.5 *	8.9	2.1
Obesity:				
Revisions	W = 0.5 *	D = 0.5 *	43.4	5.0
Editors	W = 0.7 *	D = 0.5 *	13.1	2.8
Cancer:				
Revisions	W = 0.6 *	D = 0.5 *	21.3	3.4
Editors	W = 0.6 *	D = 0.5 *	13.4	2.9
Asthma:				
Revisions	W = 0.4 *	D = 0.5 *	236.1	11.6
Editors	W = 0.4 *	D = 0.5 *	604.4	19.2
Cardiovascular Diseases:				
Revisions	W = 0.5 *	D = 0.5 *	55.2	5.7
Editors	W = 0.7 *	D = 0.5 *	17.9	2.8

TABLE 6.2: Results of Normality Test Conducted on Each of the Main Article's Number of Revisions Per Day and Number of Editors Per Day (*p $\leq$ 0.05)

6.3   Assessment of Outlier Detection Methods

As a prior step to answering Q1 and Q2, it was crucial to assess the different methods of outlier detection to 1) detect and compare the percentage of anomalies found in the data set of the number of revisions per day or in the data set of number of unique editors per day; 2) select the method that differentiate between editing activities on Wikipedia articles related to new disease outbreak crises and editing activities on Wikipedia articles related to chronic diseases; 3) select the most accurate method to measure burstiness in articles related to new disease outbreaks and in articles related to chronic diseases. This was done by finding the outliers in the dataset of the number of revisions per day of the Zika virus article. These outliers were then compared with the outliers labeled in the ground-truth data. The accuracy was then measured on the basis of equation 5.15 to find an outlier detection method that produced the most accurate result. The accuracy of different outlier detection methods is presented in 6.3. Table 6.3 affirms that

Method	Accuracy
Z-score	88.8 %
Tukey's method	94.6%
Local Outlier Factor (LOF)	98.8%
Seasonal Hybrid ESD (S-H-ESD)	96.0%
Breakout Detection	84.8%
ARIMA(1,0,0)	91.3%
ARIMA(0,1,0)	89.9%
ARIMA(1,1,0)	90.6%
ARIMA(0,1,1)	90.3%
ARIMA(0,2,1) and ARIMA(0,2,2)	90.6%
ARIMA(1,1,2)	90.3%

TABLE 6.3: Assessment of Different Outlier Detection Methods

the accuracy of Tukey's Method, Local Outlier Factor (LOF), Seasonal Hybrid ESD (S-H-ESD), and ARIMA (1,0,0) models was comparable. These four models achieved an accuracy of more than 90%. Thus, these four models were selected to detect the outliers in articles related to new disease outbreaks and chronic diseases.

6.4   Anomalies in Wikipedia Articles

Anomaly is the percentage of outliers that are detected in *either* he data set of the number of revisions per day *or* the data set of number of unique editors per day. The anomalies that were detected in the data sets of the number of revisions per day were reported in Table 6.4 and Table 6.6, while the anomalies that were detected in the data set of the number of editors per day for each of the main articles were reported in Table 6.5 and Table 6.7. Different outlier detection models yielded different results. The

ARIMA model was more sensitive to outlier detection than the Tukey's method, LOF and S-H-ESD models. This was because the ARIMA model detected a higher percentage of outliers (anomalies) than the other models. For example, the outlier percentage that was detected by the ARIMA model was 12.9% for the Ebola virus disease article, while the outlier percentage that was detected by the S-H-ESD model was only 1.97% for the same article. Overall, the ARIMA model detected a higher percentage of outliers or anomalies in the different data sets of the number of revisions and the number of editors in articles related to new disease outbreaks and articles related to chronic diseases. The overall percentage of outliers (anomalies) found in the number of revisions in articles related to new disease outbreaks was comparable to the overall percentage of outliers found in articles related to chronic diseases when applying the Tukey's method, S-H-ESD and ARIMA models. Likewise, the overall percentage of outliers (anomalies) found in the number of editors in articles related to new disease outbreaks was comparable to the overall percentage of outliers found in articles related to chronic diseases when applying the Tukey's Method, S-H-ESD and ARIMA models. Nevertheless, the overall percentage of outliers detected in the different data set of number of revisions as well as the data set of the number of editors in articles related to new disease outbreaks was slightly higher than the overall percentage of outliers detected in articles related to chronic diseases in the case of the LOF model. This could suggest that patterns in articles related to new disease outbreaks might not conform to the expected or stable editing activities. This implied that articles related to new disease outbreaks could receive sudden attention by editors. This motivated us to further investigate this deviation by measuring the burstiness in editing activities. Given that the LOF model can highly differentiate between editing activities on Wikipedia articles related to new disease outbreak crises and editing activities on Wikipedia articles related to chronic diseases as well as the LOF model has the highest accuracy rate among all the presented outlier detection methods (Table 6.3), further research was carried out to detect and compare the burstiness in articles related to new disease outbreaks and articles related chronic diseases by applying the LOF model.

6.4.1 New Disease Outbreaks

Articles	Tukey's Method	LOF	ESD (S-H-ESD)	ARIMA
Zika Virus	9.1 %	13.4 %	1.8 %	11.6 %
Ebola Virus Disease	6.4%	2.6 %	1.97 %	12.9 %
MERS	6.4%	7.5%	1.5%	14.3%
H7N9	7.7 *%	10.6%	2%	14.8%
Swine Influenza	10.3%	4.7%	1.4%	15.9%
Overall	7.4%	4.7%	1.8%	13.7%

TABLE 6.4: Percentage of Outliers Detected in the Number of Revisions Per Day

Articles	Tukey's Method	LOF	ESD (S-H-ESD)	ARIMA
Zika Virus	6.1 %	7.2 %	1.8 %	4.7 %
Ebola Virus Disease	7.0%	0.93 %	1.97 %	7.3 %
MERS	9.4%	5.2%	1.5%	4.5%
H7N9	6.3%	19.1%	1.4%	9.1%
Swine Influenza	14.1%	2.9%	1.4%	14.1%
Overall	8.6%	2.6%	1.8%	8.4%

TABLE 6.5: Percentage of Outliers Detected in the Number of Editors Per Day (Including Bots and Anonymous Editors)

6.4.2 Chronic Diseases

Articles	Tukey's Method	LOF	ESD (S-H-ESD)	ARIMA
Diabetes	5.1 %	1.7 %	1.7 %	11.5 %
Obesity	9.1%	3.1 %	1.6 %	15.1 %
Cancer	9.6%	1.9%	1.6%	13.1%
Asthma	9.0%	1.8%	1.6%	13.1%
Cardiovascular Disease	5.2%	2.4%	1.5%	7.1%
Overall	7.7%	2.5%	1.6%	12.3%

TABLE 6.6: Percentage of Outliers Detected in the Number of Revisions Per Day

Articles	Tukey's Method	LOF	ESD (S-H-ESD)	ARIMA
Diabetes	2.8 %	0.65 %	1.7 %	5.9 %
Obesity	9.8%	0.69 %	1.6 %	13.3 %
Cancer	6.6%	0.74%	1.6%	9.6%
Asthma	3.4%	0.39%	0.94%	6.5%
Cardiovascular Disease	6%	0.65%	0.28%	4.3%
Overall	5.5%	0.6%	1.3%	8.4%

TABLE 6.7: Percentage of Outliers Detected in the Number of Editors Per Day (Including Bots and Anonymous Editors)

6.5   Burstiness

In the previous chapter (Chapter 5), we argued that burstiness is a sudden increase in an activity for a short period of time. Correspondingly, burstiness was measured in this thesis as a surge in both the number of revisions and the number of unique editors over time. Burstiness is the percentage of outliers that are *simultaneously* detected in both the data set of the number of revisions per day *and* the number of unique editors per day. Incorporating the daily number of unique editors—excluding bots—into the LOF model helps to distinguish between the spike of edits that occurs as a result of one or two editors contributing to articles and the spike that occurs as a result of many editors contributing to articles collectively. Figures 6.2 to 6.6 show the burstiness in articles related to new disease outbreaks, while Figures 6.7 to 6.11 show the burstiness in articles related to chronic diseases. These figures show all observations from 2001 till 2016. The results showed that articles related to chronic diseases experienced either no or a few bursty periods ($\leq$ three instances of burstiness in total). For example, the cardiovascular disease article had two instances of burstiness in total. Moreover, the diabetes article experienced no burstiness in editing activities, as shown in Figure 6.7. This was because the outliers were not detected simultaneously in the the datasets of the number of revisions and the number of editors per day. In contrast, articles related to new disease outbreaks experienced more bursty periods, often during the outbreaks. For example, the Zika article experienced more than 12 instances of burstiness, as shown in Figure 6.2. Furthermore, the swine influenza article experienced more than nine instances of burstiness, as shown in Figure 6.6. This was because the outliers were detected simultaneously in the the datasets of the number of revisions and the number of editors per day. This implied that there were differences in the level of burstiness between the articles related to new disease outbreaks and the articles related to chronic diseases. The next section looks at classifying and comparing different levels of burstiness by applying the head/tail breaks classification scheme. This classification helps to distinguish the patterns of editing behaviours in articles related to new disease outbreaks from those in articles related to chronic diseases.
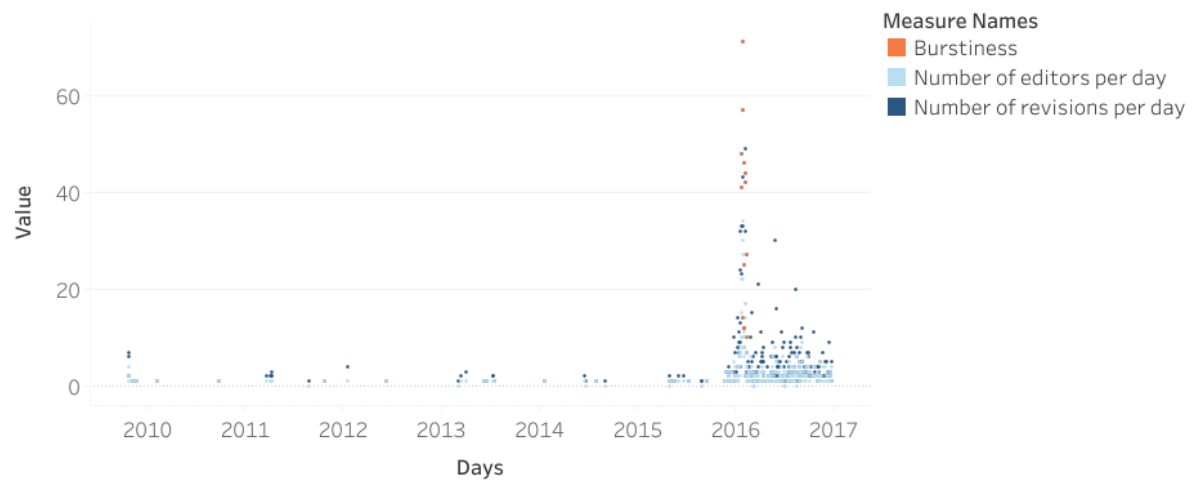
FIGURE 6.2: Burstiness in the Zika Article (New Disease Outbreaks)
https:
//public.tableau.com/views/ZikaBurstinessThesis_corr-1/Dashboard2?:
embed=y&:display_count=yes&publish=yes&:origin=viz_share_link
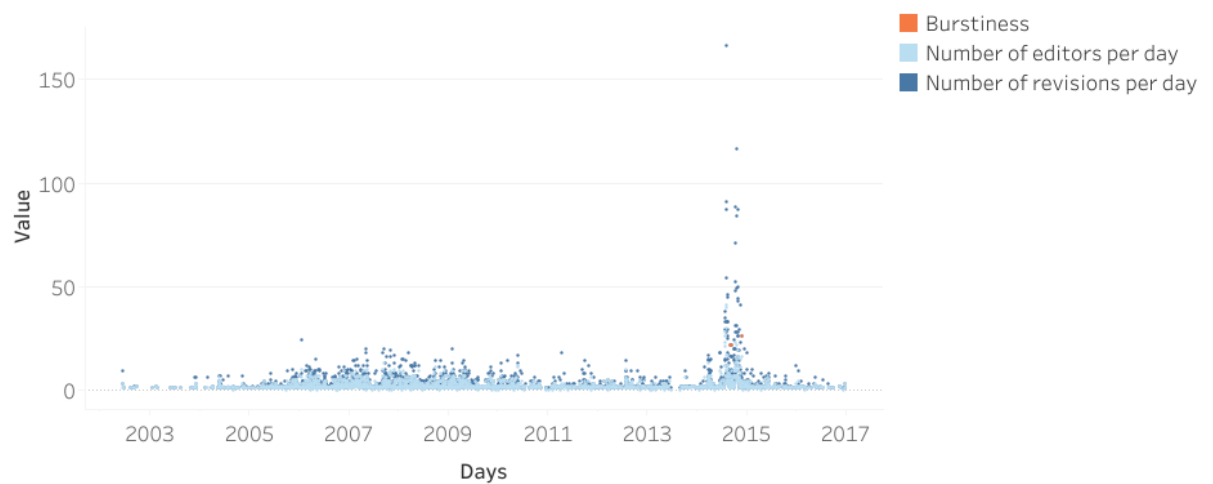
FIGURE 6.3: Burstiness in the Ebola Article (New Disease Outbreaks)
https://public.tableau.com/views/EbolaVirusThesis2/Dashboard2?:embed=y&:
display_count=yes&publish=yes&:origin=viz_share_link
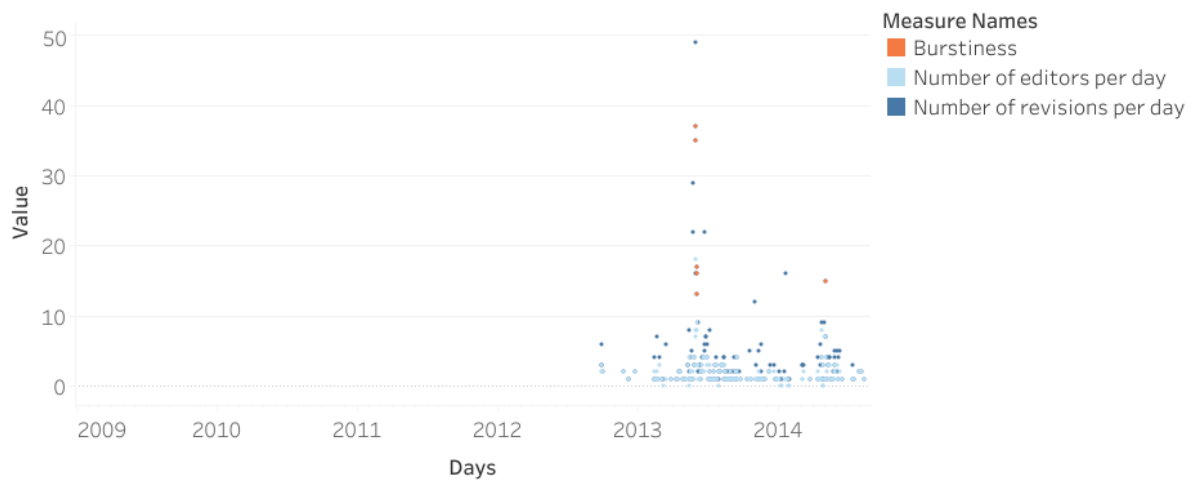
FIGURE 6.4: Burstiness in the MERS Article (New Disease Outbreaks)
https://public.tableau.com/views/MERSThesis2/Dashboard2?:embed=y&:
display_count=yes&publish=yes&:origin=viz_share_link
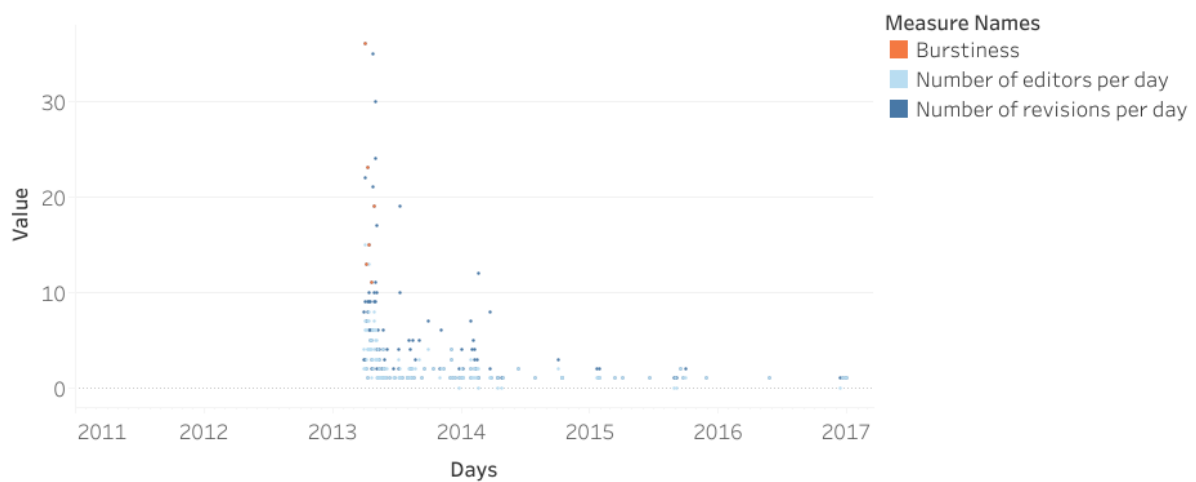
FIGURE 6.5: Burstiness in the H7N9 Article (New Disease Outbreaks)
https://public.tableau.com/views/H7N9Thesis2/Dashboard2?:embed=y&:
display_count=yes&:origin=viz_share_link
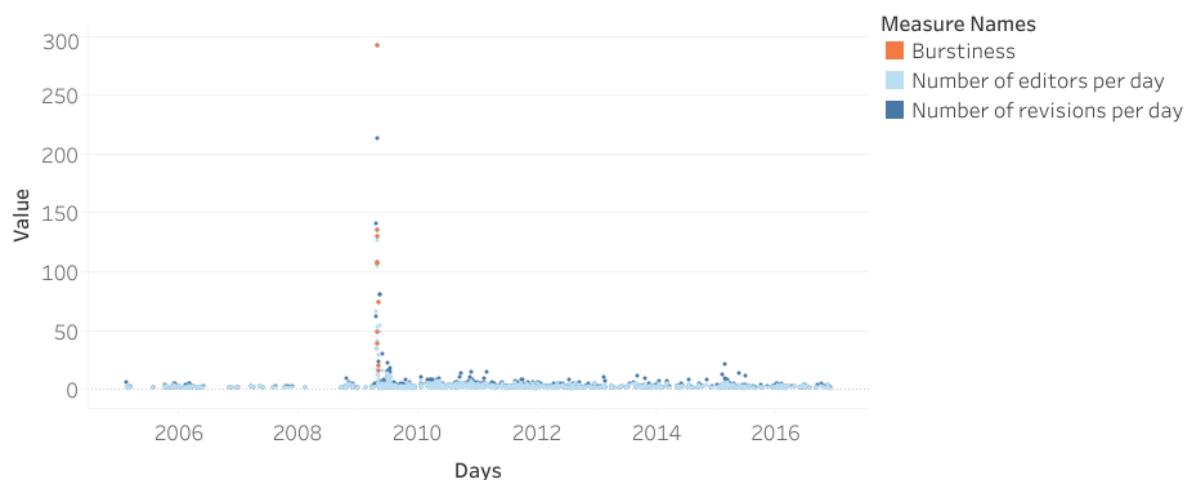
FIGURE 6.6: Burstiness in the Swine Influenza Article (New Disease Outbreaks)
https://public.tableau.com/views/SwineInfluenzaThesis2/Dashboard2?:
embed=y&:display_count=yes&publish=yes&:origin=viz_share_link
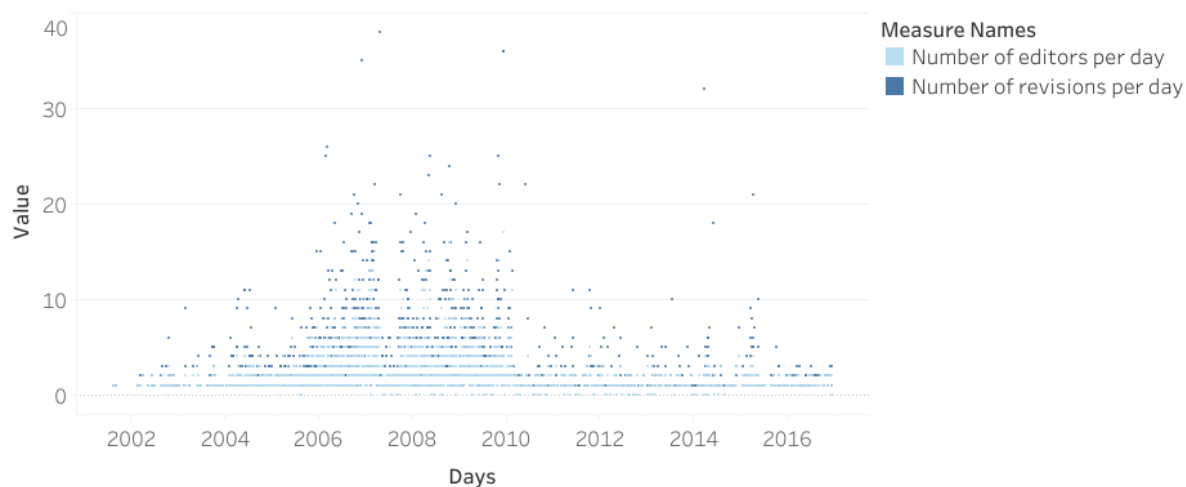
FIGURE 6.7: Burstiness in the Diabetes Article (Chronic Diseases)
https://public.tableau.com/views/DiabetesThesis2/Dashboard1?:embed=y&:
display_count=yes&publish=yes&:origin=viz_share_link
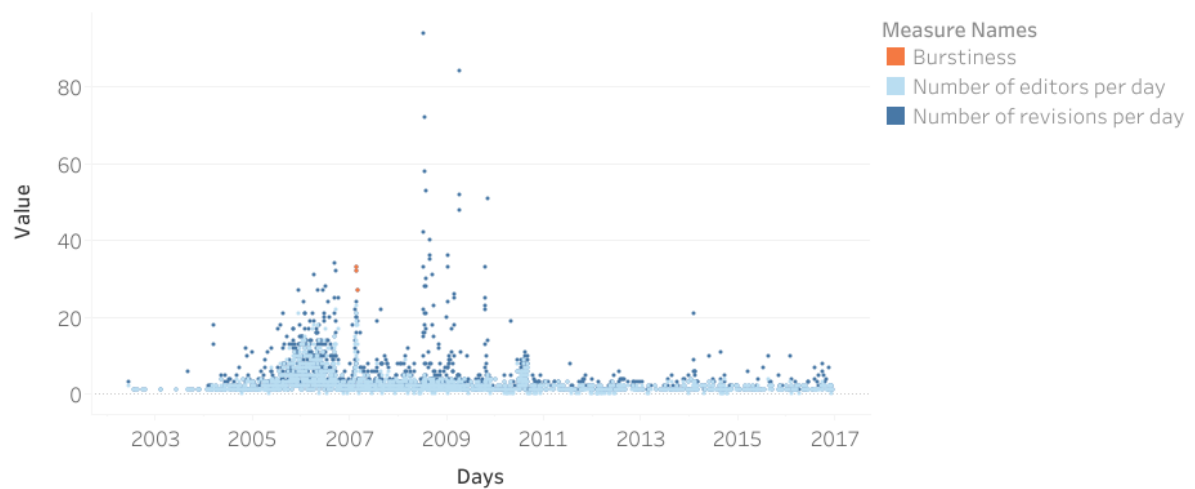
FIGURE 6.8: Burstiness in the Obesity Article (Chronic Diseases)
https://public.tableau.com/views/ObesityThesis2/Dashboard1?:embed=y&:
display_count=yes&publish=yes&:origin=viz_share_link
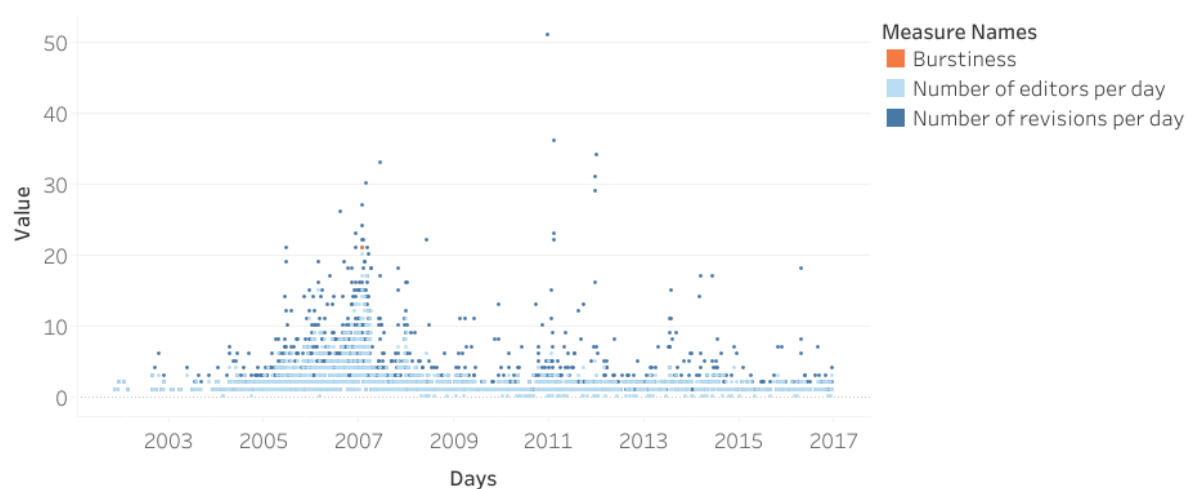
FIGURE 6.9: Burstiness in the Cancer Article (Chronic Diseases)
https://public.tableau.com/views/CancerThesis2/Dashboard1?:embed=y&:
display_count=yes&publish=yes&:origin=viz_share_link
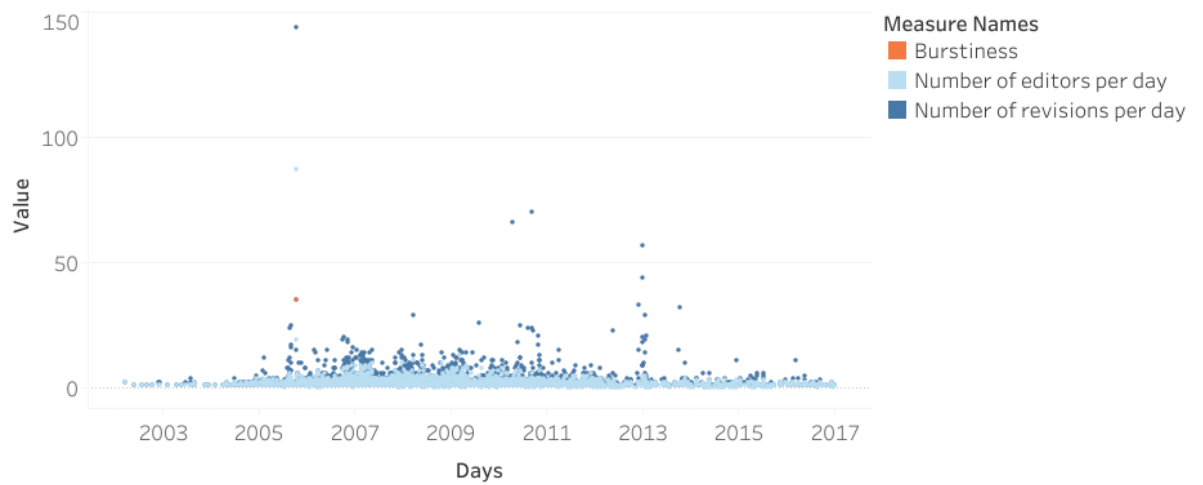
FIGURE 6.10: Burstiness in the Asthma Article (Chronic Diseases)
https://public.tableau.com/views/AsthmaThesis2/Dashboard1?:embed=y&:
display_count=yes&publish=yes&:origin=viz_share_link
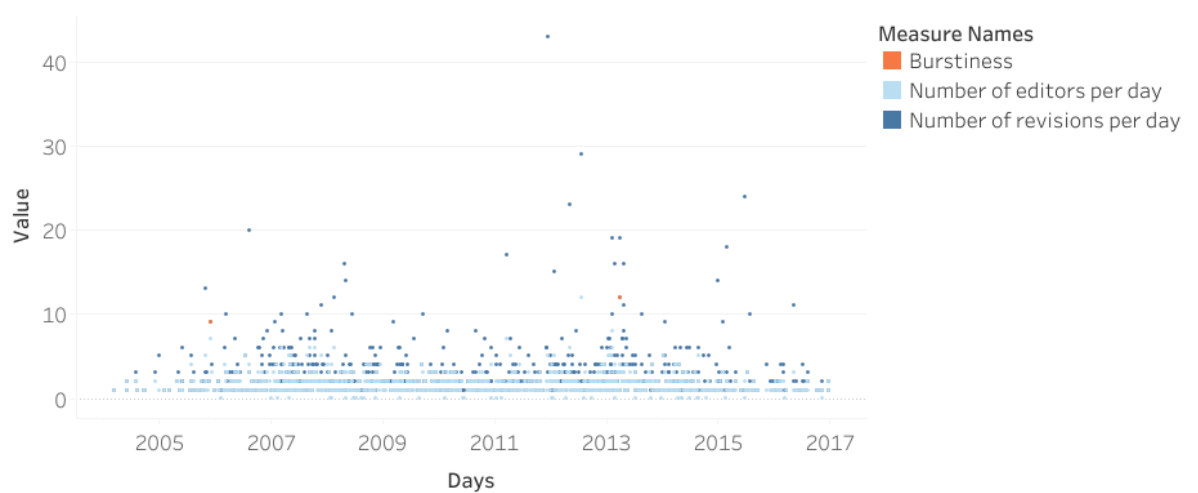
FIGURE 6.11: Burstiness in the Cardiovascular Disease Article (Chronic Diseases)
https://public.tableau.com/views/HeartDiseasesThesis2/Dashboard1?:
embed=y&:display_count=yes&publish=yes&:origin=viz_share_link

6.6   Levels of Burstiness

In this section, the levels of burstiness were categorised as high, moderate, or low. These categories helped to compare between patterns of editing activities in articles related to new disease outbreaks and patterns of editing activities in articles related to chronic diseases. Moreover, these categories helped to compare patterns of editing activities across articles related to new disease outbreaks and across articles related to chronic diseases.

Before categorising the levels of burstiness and applying the head/tail Breaks classification scheme, it was important to test whether the distribution of all the editing activities followed a heavy-tailed distribution. Figures 6.12 and 6.13 show the distribution of the number of edits (revisions) and the daily number of unique editors for all the selected articles related to new disease outbreaks. Further, Figures 6.16 and 6.17 show the distribution of the number of edits (revisions) and the daily number of unique editors for all the selected articles related to chronic diseases. The distribution of both the number of revisions and the number of unique editors was skewed right as most of the articles had a small daily number of revisions and a small daily number of unique editors. Figures 6.14, 6.15, 6.18, and 6.19 show the plots of the distributions along the log-log axis. The distribution line was straight when graphed as a double-logarithm plot. This straight distribution line, which is the hallmark of a power law, is often used to detect the power law in a double-logarithm plot (Jiang, 2013).

FIGURE 6.12: Distribution of the Number of Revisions (New Disease Outbreaks)

FIGURE 6.13: Distribution of the Number of Editors (New Disease Outbreaks)
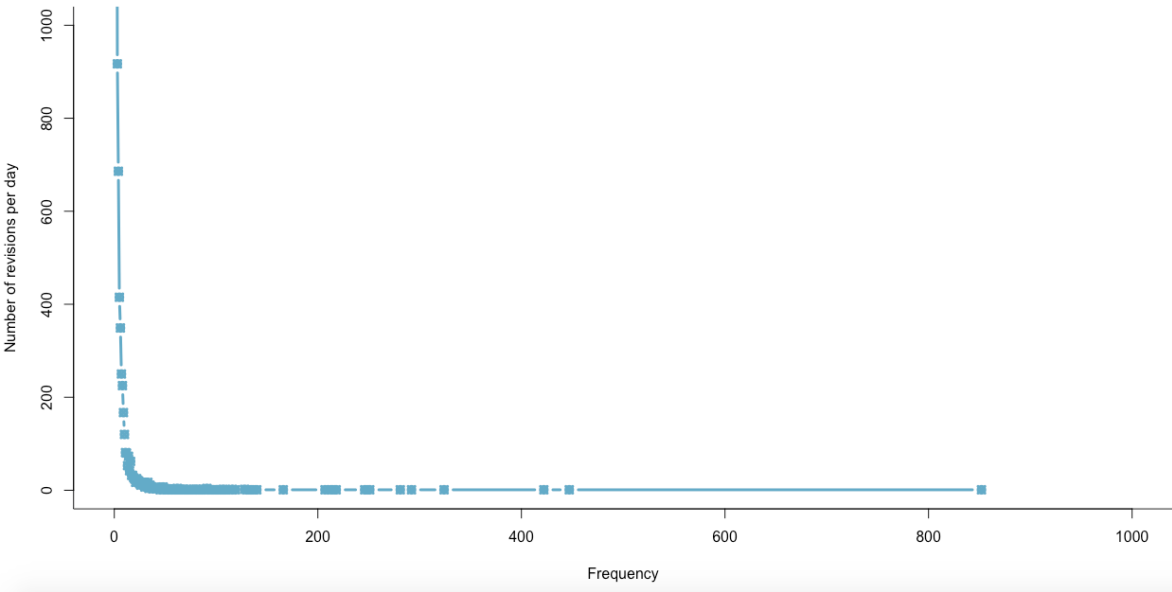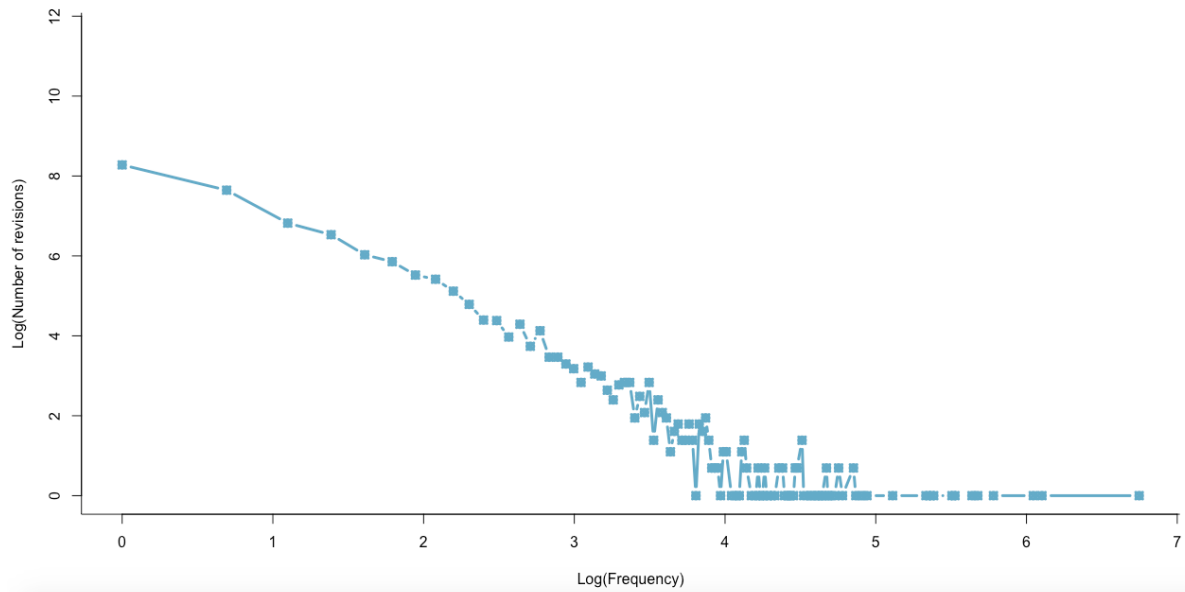
FIGURE 6.14: Distribution of the Number of Revisions (Log Scale - New Disease Outbreaks)
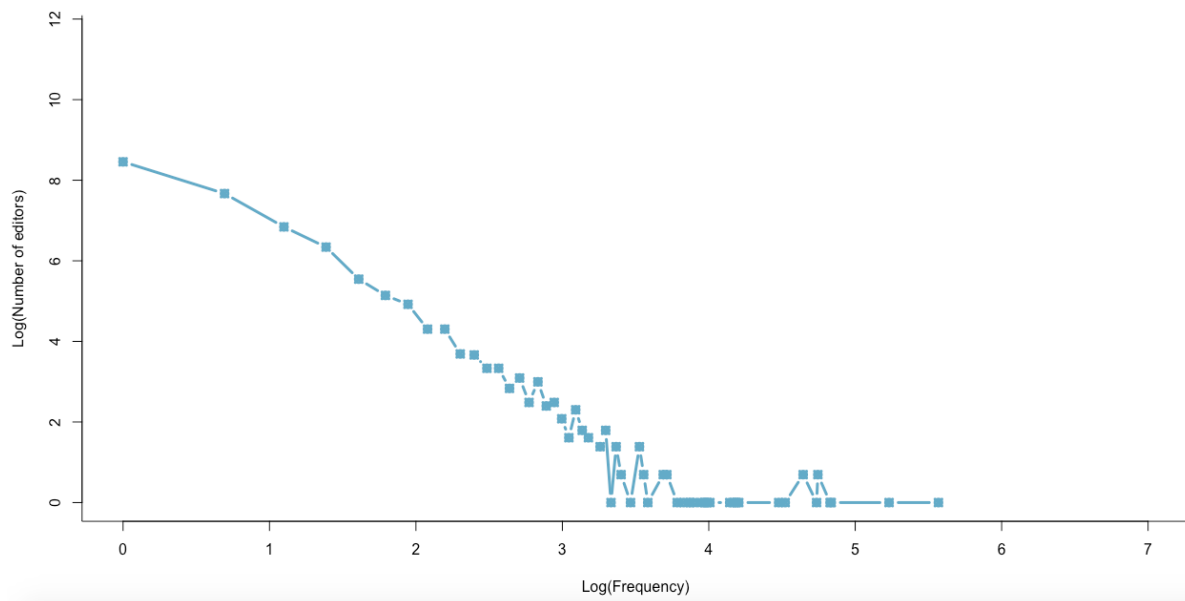
FIGURE 6.15: Distribution of the Number of Editors (Log Scale - New Disease Outbreaks)
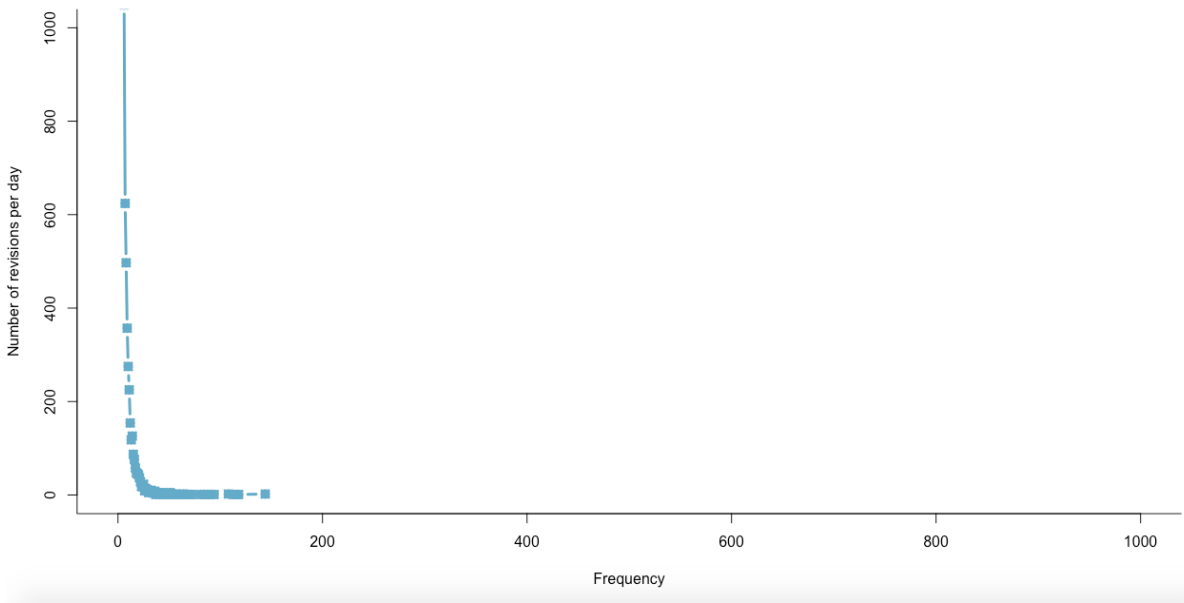
FIGURE 6.16: Distribution of the Number of Revisions (Chronic Diseases)
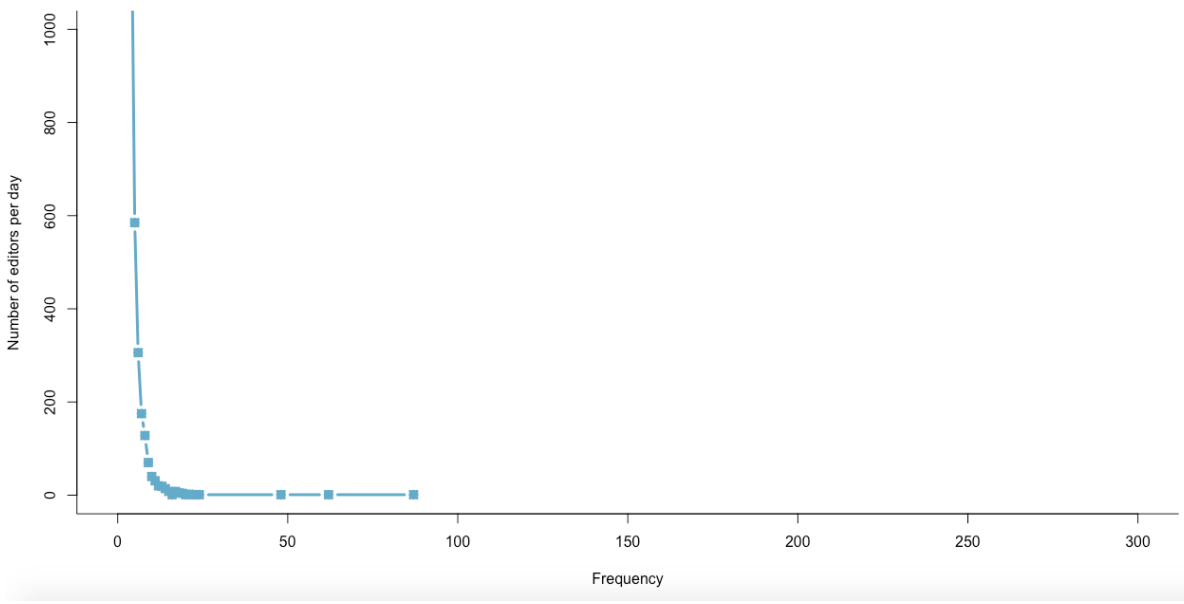
FIGURE 6.17: Distribution of the Number of Editors (Chronic Diseases)
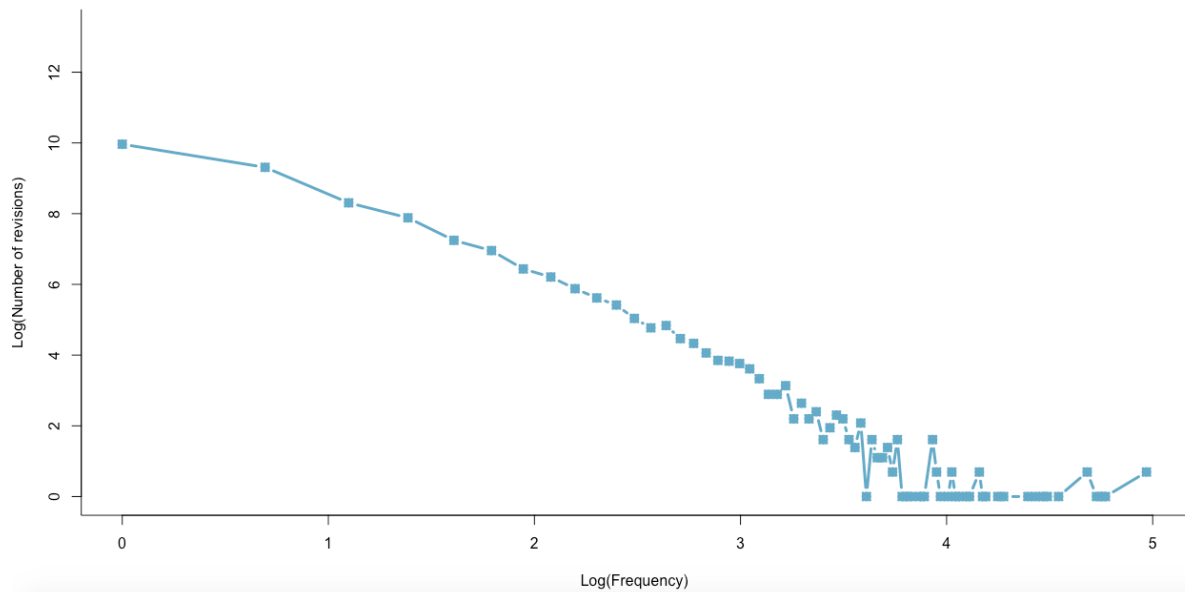
FIGURE 6.18: Distribution of the Number of Revisions (Log Scale - Chronic Diseases)
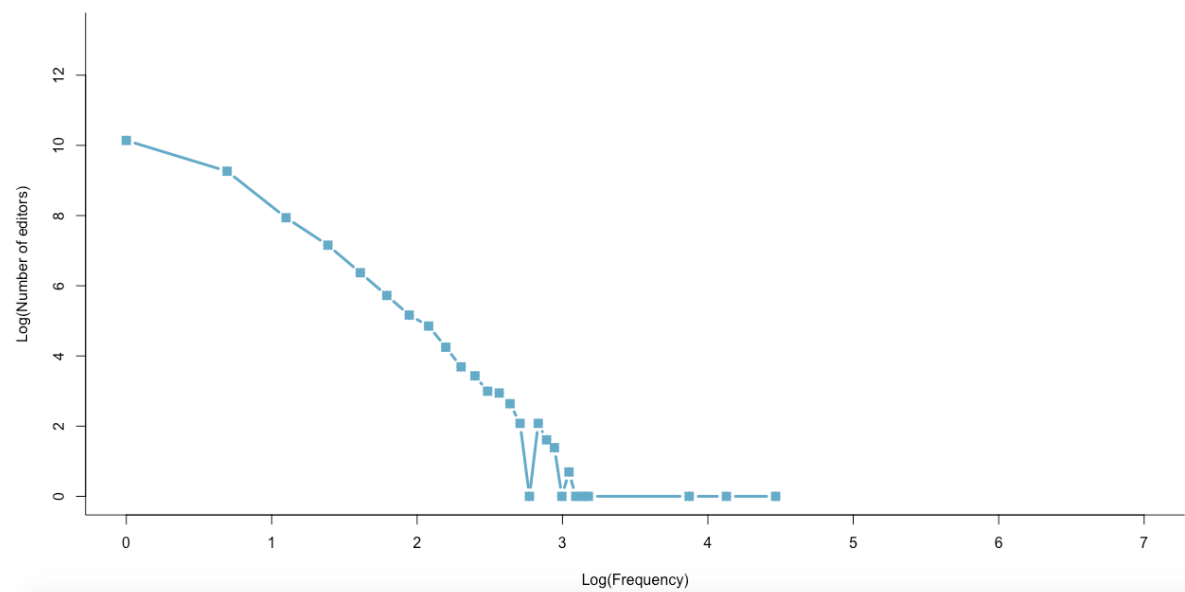
FIGURE 6.19: Distribution of the Number of Editors (Log Scale - Chronic Diseases)

Based on the head/tail breaks classification scheme for data with a heavy-tailed distribution, the burstiness was classified in the articles related to new disease outbreaks (Table 6.8) and chronic diseases (Table 6.9). Different thresholds of classifications emerged in articles related to new disease outbreaks and chronic diseases. In order for a Wikipedia article related to a new disease outbreak to be classified as having high burstiness, the number of revisions should be greater than or equal to 90 and the number of unique editors should be greater than or equal to 33. Moreover, in order for an article to be classified as having moderate burstiness, the number of revisions should be greater than or equal to 39 and the number of unique editors should be greater than or equal to 15. Moreover, in order for an article to be classified as having low burstiness, the number of revisions should be greater than or equal to 15 and the number of unique editors should be greater than or equal to 6. Alternatively, in order for a Wikipedia article related to chronic diseases to be classified as having high burstiness, the number of revisions should be greater than or equal to 17 and the number of unique editors should be greater than or equal to 5. Furthermore, in order for an article to be classified as having moderate burstiness, the number of revisions should be greater than or equal to 12 and the number of unique editors should be greater than or equal to 4. Finally, in order for an article to be classified as having low burstiness, the number of revisions should be greater than or equal to 10 and the number of unique editors should be greater than or equal to 3. These findings proved that articles related to new disease outbreaks experienced a higher degree of burstiness than articles related to chronic diseases. These findings drove us to investigate thoroughly the phenomenon of burstiness in articles related to new disease outbreaks.

Category	Number of Revisions	Number of Unique Editors
High	≥ 90	≥ 33
Medium	≥ 39	≥ 15
Low	≥ 15	≥ 6

TABLE 6.8: New Disease Outbreak Classifications

Category	Number of Revisions	Number of Unique Editors
High	≥ 17	≥ 5
Medium	≥ 12	≥ 4
Low	≥ 10	≥ 3

TABLE 6.9: Chronic Disease Classifications

6.7   Summary

The goal of this chapter is to disclose the characteristics of Wikipedia editorial activities during new disease outbreak crises (RQ1) and compare editorial activities on Wikipedia articles related to new disease outbreak crises and editorial activities on Wikipedia articles related to chronic diseases (RQ2). These questions were answered by applying outlier detection methods and head/tail breaks classification scheme to observe two different phenomena: anomaly and burstiness. Anomaly is the percentage of outliers— points that diverge from the majority of observations in a data set—that are detected in either the data set of the number of revisions per day or the data set of number of unique editors per day. However, burstiness is the percentage of outliers that are simultaneously detected in both the data set of the number of revisions per day and the number of unique editors per day.

This chapter started by demonstrating that the number of edits and the number of editors of articles in our sample were not normally distributed and were skewed to the right. This suggested that outliers could exist in the articles considered in our sample. Therefore, different outlier detection methods were assessed and compared for three different reasons.

The first reason is to detect and compare the percentage of outliers or anomalies found in the data set of the number of revisions per day or in the data set of number of unique editors per day. The results suggested that the percentage of outliers (anomalies) in articles related to new disease outbreaks was higher than percentage of outliers in articles related to chronic diseases.

The second reason is to select the outlier detection model that differentiates between editing activities on Wikipedia articles related to new disease outbreak crises and editing activities on Wikipedia articles related to chronic diseases. The results divulged that the LOF model can highly differentiate between editing activities on Wikipedia articles related to new disease outbreak crises and editing activities on Wikipedia articles related to chronic diseases. This is because the overall percentage of outliers (anomalies) detected in the distinct data set of number of revisions as well as the data set of the number of editors in articles related to new disease outbreaks was slightly higher than the overall percentage of outliers detected in articles related to chronic diseases when applying the LOF model.

The third reason is to select the most accurate method to measure burstiness in articles related to new disease outbreaks and in articles related to chronic diseases. The results affirmed that LOF model had the highest accuracy rate among all the outlier detection methods. Accordingly, the phenomenon of burstiness was explored by applying

the LOF model. Extending this study to explore the phenomenon of burstiness is important to secernate between the spike of edits that occurs as a result of one or two editors contributing to articles and the spike that occurs as a result of editors contributing to articles conjointly. This chapter confirmed that articles related to new disease outbreaks experienced more bursty periods than articles related to chronic diseases. After applying the head/tail breaks classification scheme, this chapter also confirmed that articles related to new disease outbreaks experienced a higher degree of burstiness than articles related to chronic diseases. Exploring and measuring the phenomenon of burstiness have helped in comparing between editors' activities and attention to articles related to new disease outbreaks and articles related to chronic diseases. As a methodological contribution, burstiness could be measured to compare and contrast editors' activities and attention to Wikipedia articles that cover similar or different topics. In addition, burstiness could be measured to compare and contrast editors' activities and attention to Wikipedia articles in different languages. The following chapter will focus on studying factors that might influence burstiness in Wikipedia during new disease outbreaks.

Chapter 7

Results: Factors Behind Burstiness in Wikipedia

'There also is the plight that comes from natural disasters; these natural disasters could be alleviated or dealt with; we only need some time to do it.'

— Bhumibol Adulyadej
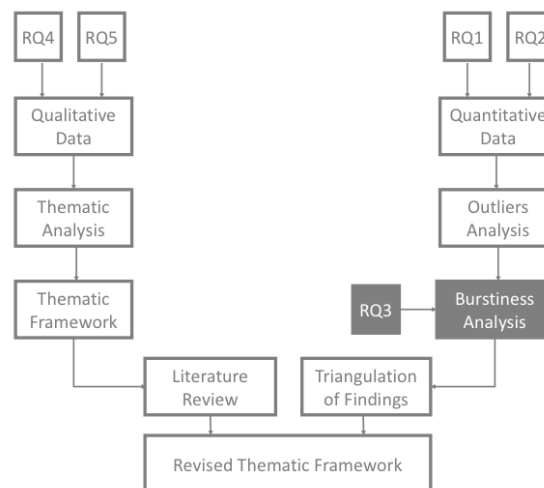
FIGURE 7.1: Chapter's Placement Within Wider Research Design

The previous chapter demonstrated that articles related to new disease outbreaks have experienced more intense burstiness than articles related to chronic diseases. The head-/tail breaks classification scheme revealed that articles related to new disease outbreaks have different and usually higher levels of burstiness than articles related to chronic diseases. This chapter builds on these results by firstly finding the disease outbreaks articles that have experienced high, medium, or low burstiness. Secondly, this chapter explores factors that might influence the level of burstiness in Wikipedia during new

disease outbreaks. In response to RQ3, factors related to news coverage, start and end of diseases, and warnings issued about new disease outbreaks were investigated. This chapter also looks at the contribution of bots to Wikipedia articles during new disease outbreak crises.

7.1   Classification of Disease Outbreak Burstiness

Applying the head/tail breaks classification scheme to articles related to new disease outbreaks revealed that articles related to Zika, Ebola, and swine influenza have experienced a high to moderate level of burstiness (Table 7.1). In spite of this, most of the articles related to MERS and H7N9 have experienced a low level of burstiness. This indicates that articles related to new disease outbreaks experience different levels of burstiness. The following section examines factors that might influence the level of burstiness in Wikipedia articles during new disease outbreak crises. The news coverage of disease outbreaks is the first factor that will be examined in this section.

Category	Article
High	Ebola virus disease
	Ebola virus cases in the United States
	2009 flu pandemic by country
	2009 flu pandemic
	Swine influenza
Moderate	Zika virus
	2015—16 Zika virus epidemic
	2009 flu pandemic in the UK
	2009 flu pandemic in Canada
Low	Zika virus outbreak timeline
	Zika fever
	Middle East respiratory syndrome
	2015 MERS outbreak in South Korea
	Influenza A virus subtype H7N9
	Avian influenza
	2009 flu pandemic in Spain
	2009 flu pandemic in Australia
	2009 flu pandemic in Asia

TABLE 7.1: Burstiness Classifications of Wikipedia Articles Related to New Disease Outbreaks

7.2   Disease Outbreaks in the News

As the public mostly relies on press reports to find information related to new disease outbreaks, it is essential to examine the relationship between the amount of news coverage and that of editing activities on Wikipedia. Table 7.2 reports the number of news articles that cover disease outbreaks as retrieved from the NEXIS database[1]. This table shows that Zika, Ebola, and swine influenza have been widely covered in the press. To further examine the relationships between the number of news articles and that of editing activities on Wikipedia, Pearson's correlation test was applied to the data from the start of the outbreak to one year after the outbreak. The data on editing activities include the number of revisions per day and the number of unique editors per day. The results presented in Tables 7.3 and 7.4 suggest that there is a positive and moderate relationship between an increase in news coverage and an increase in the editing activities on Wikipedia during disease outbreaks. These findings motivated us to extend this analysis to explore Wikipedia editing activities over time in relationship to news coverage, start and end of outbreaks, and World Health Organization (WHO) warnings.

---

[1]http://www.nexis.com/

Disease	Frequency
Zika	50939
Ebola	290375
MERS	1762
H7N9	13481
Swine influenza	294829

TABLE 7.2: Nexis Search Results of Newspapers, Newswires, Press Releases, Web-based Publications, Industry Trade Press, News Transcripts, Blogs, Newsletters

Disease	P-value	r
Zika	< 0.05	0.6
Ebola	< 0.05	0.5
MERS	< 0.05	0.4
H7N9	< 0.05	0.6
Swine influenza	< 0.05	0.4

TABLE 7.3: Pearson's Correlation Test Between Number of Revisions and Number of News Articles

Disease	P-value	r
Zika	< 0.05	0.6
Ebola	< 0.05	0.5
MERS	< 0.05	0.5
H7N9	< 0.05	0.7
Swine influenza	< 0.05	0.4

TABLE 7.4: Pearson's Correlation Test Between Number of Unique Editors and Number of News Articles

7.3   Timeline

Burstiness over time in Wikipedia articles related to new disease outbreaks is reported in Figures 7.2 to 7.6. These figures show all observations from 2001 till 2016. The start and the end of disease outbreaks are also shown in conjunction with the Wikipedia editing activities. The subplots depict the daily number of news articles from the start of an outbreak to one year after the outbreak as retrieved from NEXIS. The results showed that the articles related to MERS and H7N9 experienced less burstiness than those related to Zika, Ebola, and swine influenza. Burstiness occurred in the Zika, Ebola and swine influenza articles after the WHO declared the epidemics as public health emergencies of international concern (PHEIC)[2].

These figures and the correlation values suggest that there is a relationship between the level of burstiness and news coverage as well as a relationship between the level of

---

[2]https://www.who.int/features/qa/emergency-committees/en/

burstiness and health organisations' warnings. This proposes that the level of burstiness is influenced by not only the news coverage of disease outbreaks, but also major health organisation' warnings that these diseases could cause major threats at the local, national, and international levels.
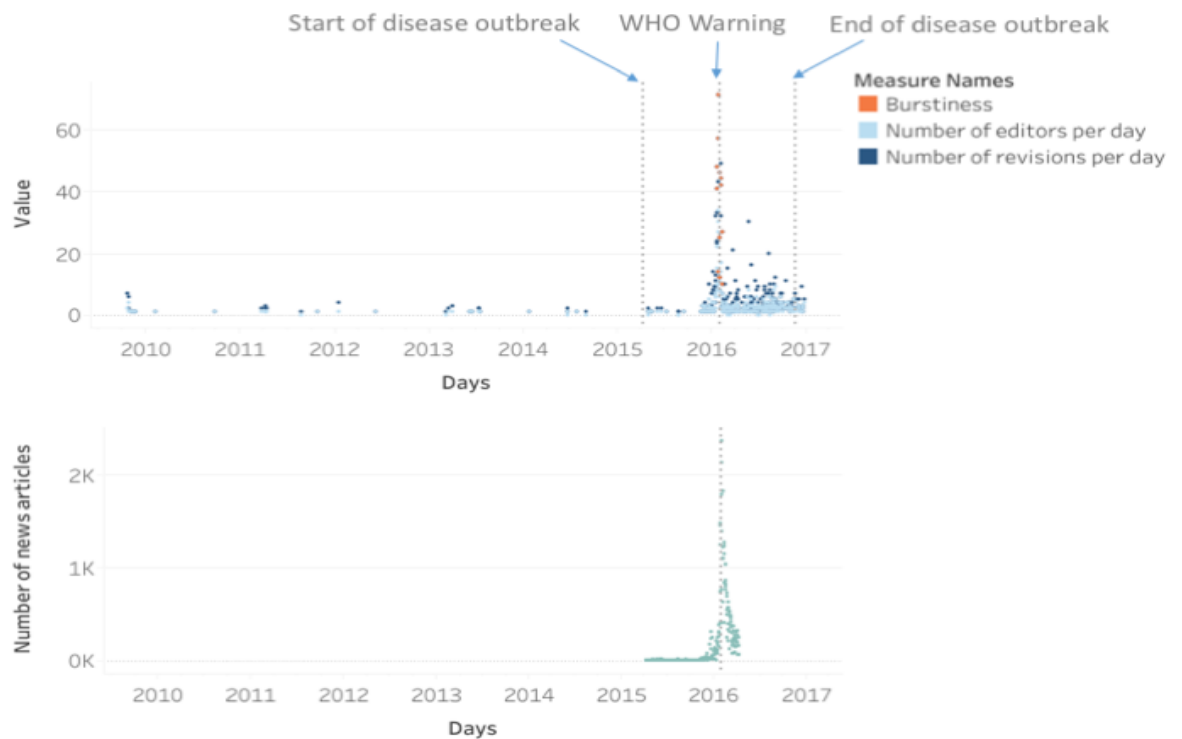
FIGURE 7.2: Zika Timeline
https://public.tableau.com/views/ZikaBurstinessThesis_timeline_corr-1/
Dashboard1?:embed=y&:display_count=yes&:origin=viz_share_link
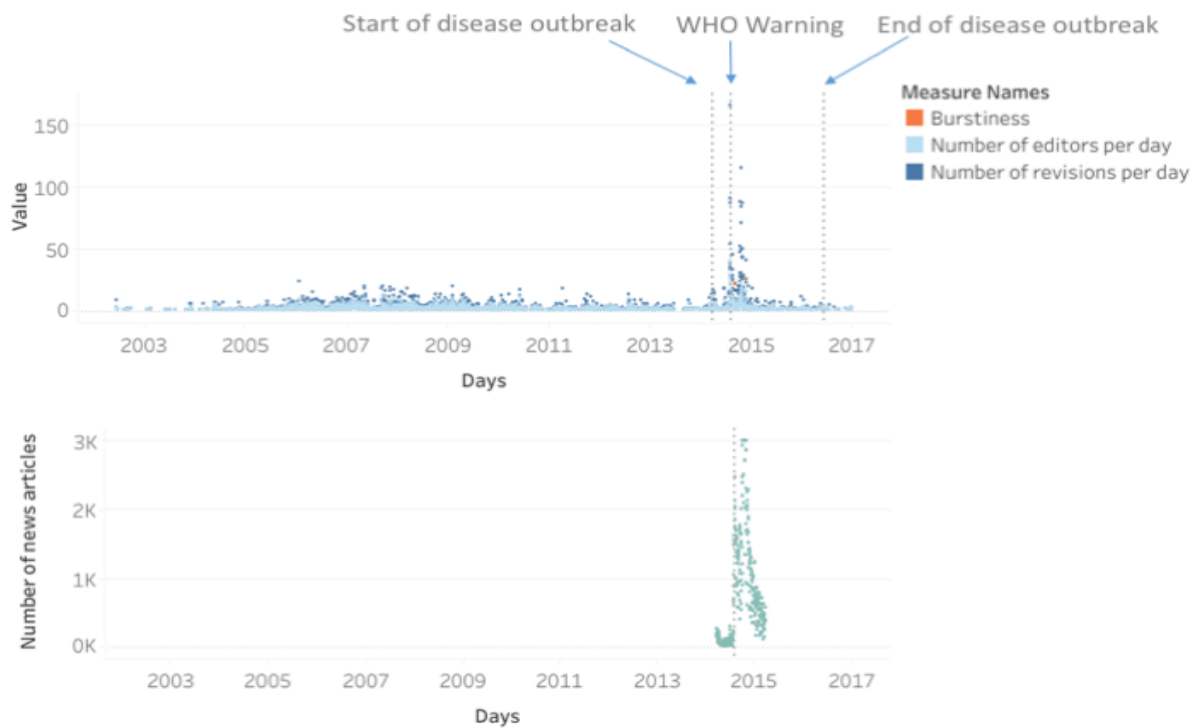
FIGURE 7.3: Ebola Timeline

https://public.tableau.com/views/EbolaVirusThesis-timeline-1-corr/
Dashboard1?:embed=y&:display_count=yes&:origin=viz_share_link
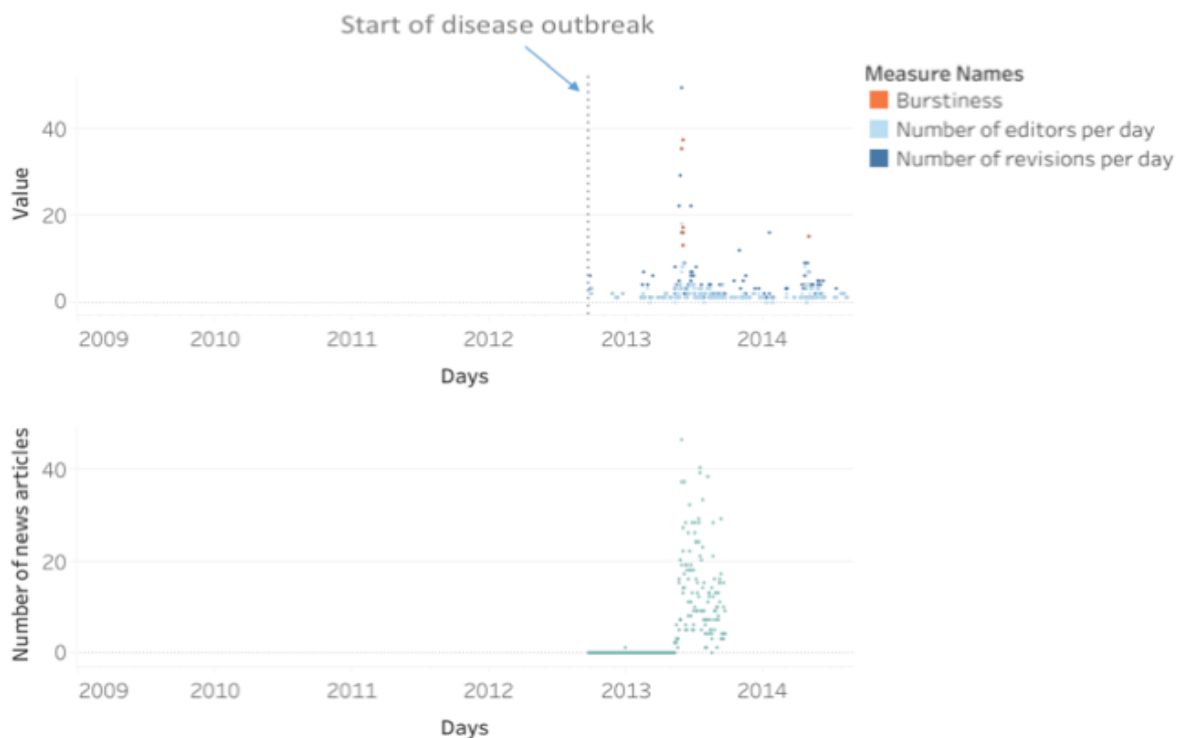
FIGURE 7.4: MERS Timeline

https://public.tableau.com/views/MERSThesisTimeline2/Dashboard1?:
embed=y&:display_count=yes&:origin=viz_share_link
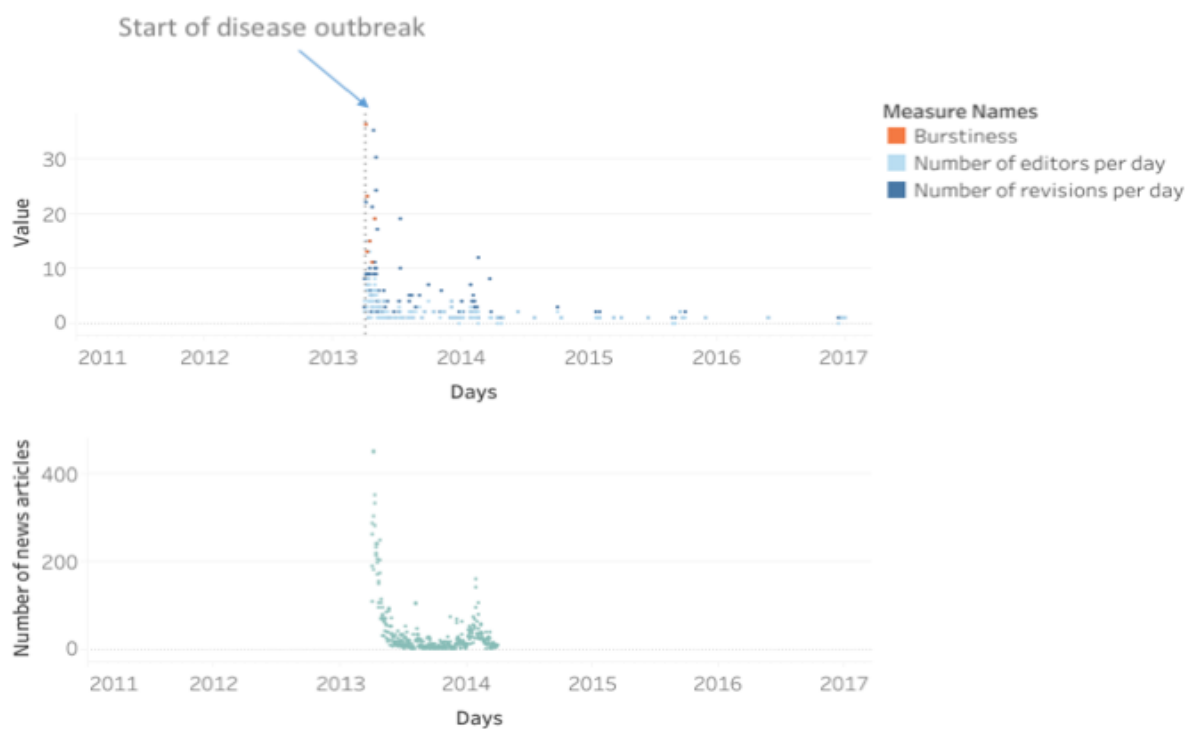
FIGURE 7.5: H7N9 Timeline
https://public.tableau.com/views/H7N9ThesisTineline2/Dashboard1?:
embed=y&:display_count=yes&:origin=viz_share_link
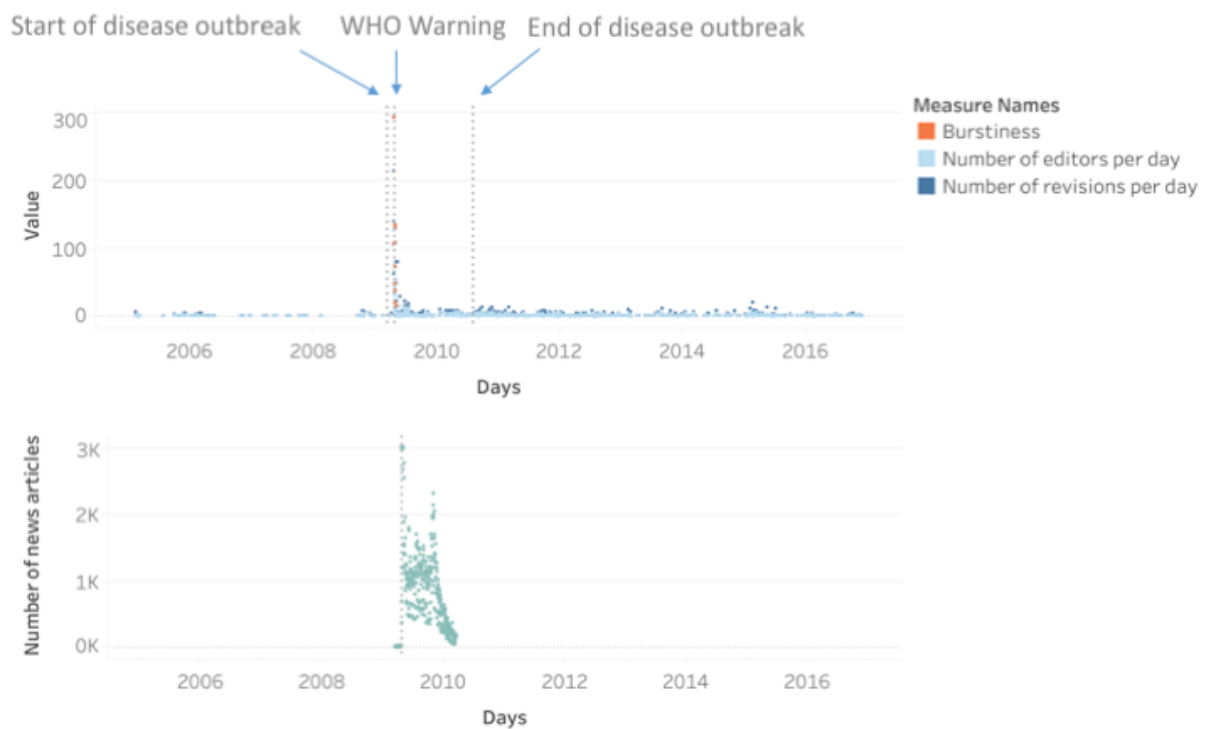
FIGURE 7.6: Swine Influenza Timeline
https://public.tableau.com/views/SwineInfluenzaTimeline2/Dashboard1?:
embed=y&:display_count=yes&:origin=viz_share_link

7.4   Bots During New Disease Outbreaks

The previous section examined burstiness as a surge in the number of revisions per day and the number of unique editors per day, but excluded the number of revisions made by bots from the analysis. Considering that Wikipedia is a socio-technological phenomenon that involves humans and bots working together to edit articles, it is important to shed light on the contribution of bots during new disease outbreaks. Thus, this section examines a surge in the number of bots' edits per day in articles related to new disease outbreaks. This was done to understand whether the surge in the number of bots' edits is another factor that might influence the level of burstiness in articles related to new disease outbreaks. Figures 7.7 to 7.11 depict the number of bots' edits over time in the Zika, Ebola, MERS, H7N9, and swine influenza articles. These figures show all observations from 2001 till 2016. Articles related to new disease outbreaks, particularly the swine influenza article, have experienced a slight increase in the number of bots' edits during the outbreaks. In fact, outliers in the datasets of the number of bots' edits per day were detected using the Local Outlier Factor model in all the five main articles related to new disease outbreaks. Still, the number of bots' edits per day identified as the outliers was smaller than the number of humans and bots' edits per day (i.e the number of revisions) identified as the outliers.
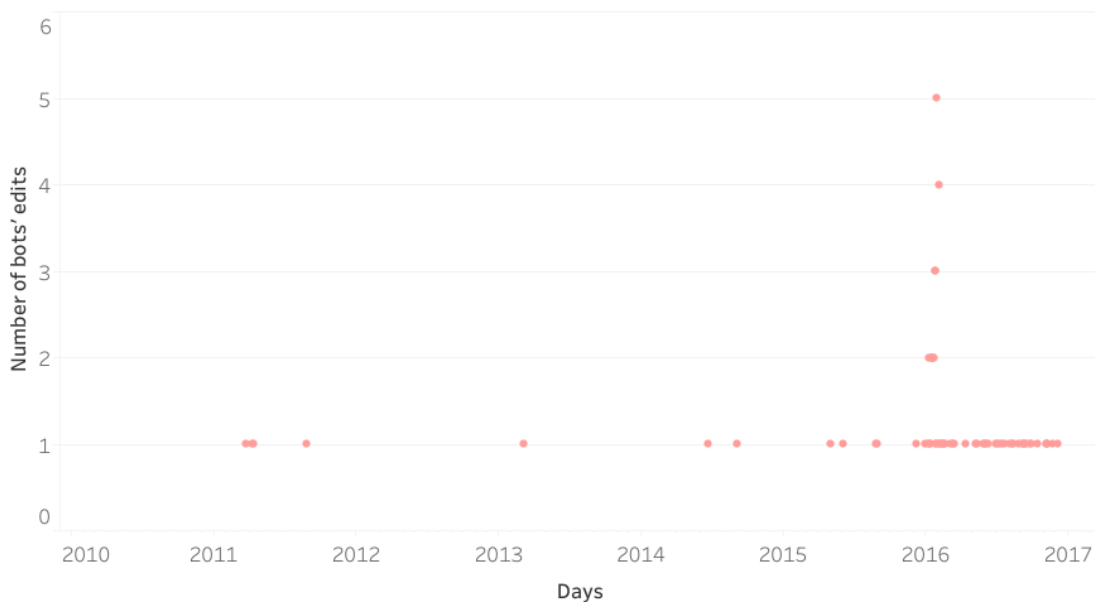
FIGURE 7.7: Bots' Edits in the Zika Article
https://public.tableau.com/views/BotsZikaVirus/Dashboard1?:embed=y&:
display_count=yes&publish=yes&:origin=viz_share_link
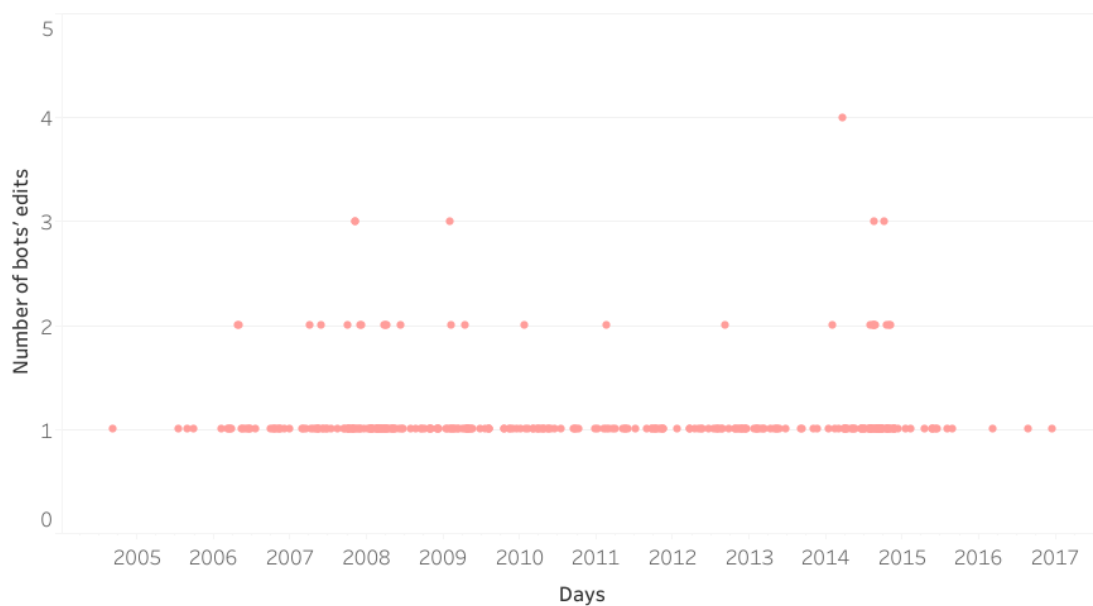
FIGURE 7.8: Bots' Edits in the Ebola Article
https://public.tableau.com/views/EbolaVirusDiseasesBots/Sheet2?:
embed=y&:display_count=yes
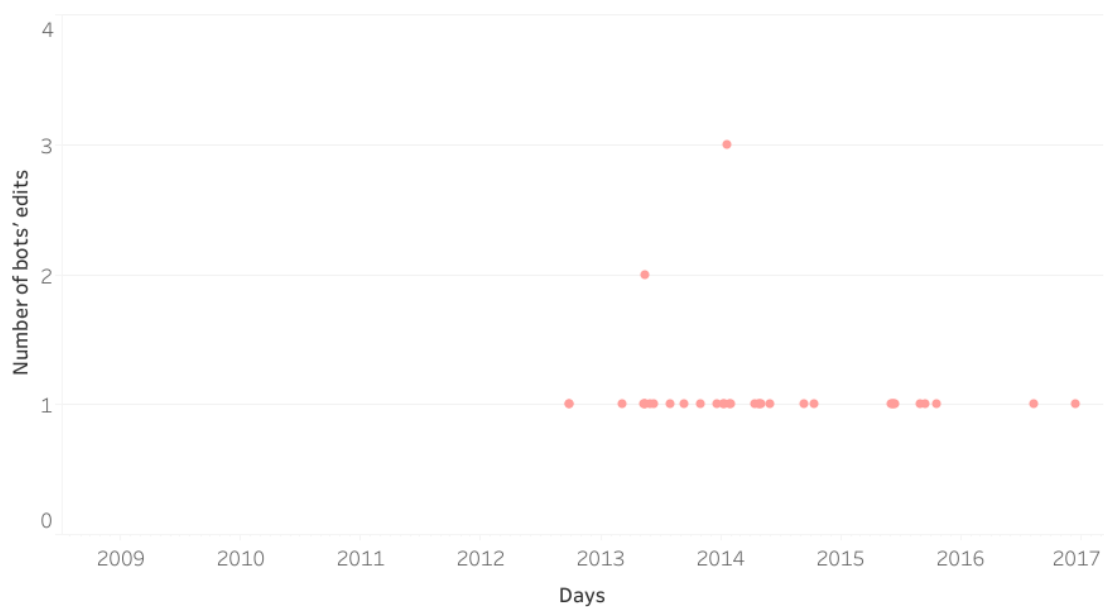
FIGURE 7.9: Bots' Edits in the MERS Article
https://public.tableau.com/views/MERSBotsedits/Sheet2?:embed=y&:
display_count=yes
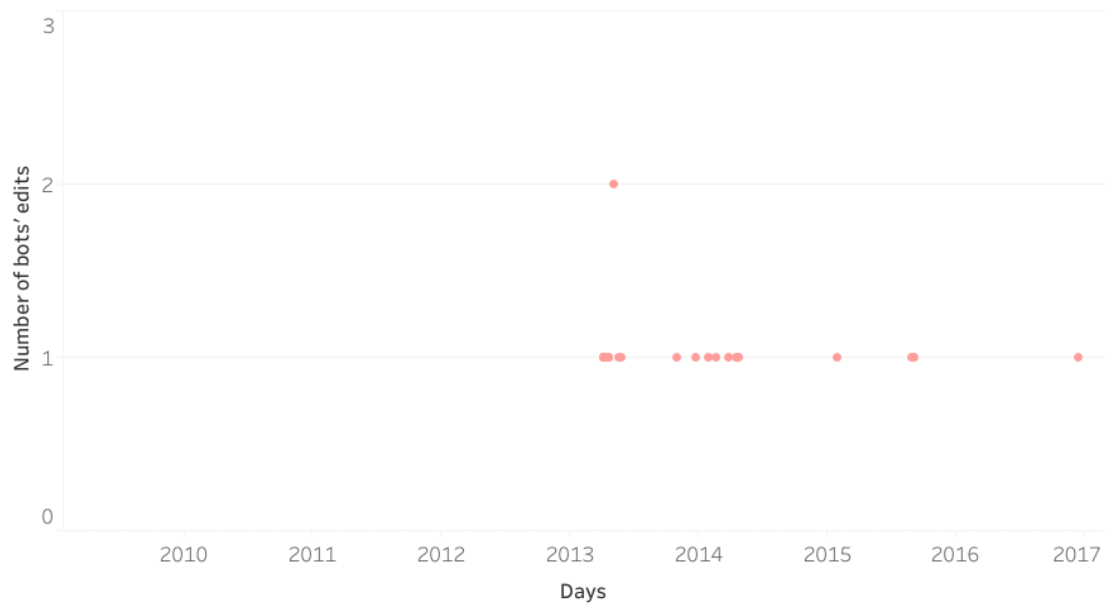
FIGURE 7.10: Bots' Edits in the H7N9 Article
https://public.tableau.com/views/H7N9Botsedits/Sheet2?:embed=y&:
display_count=yes
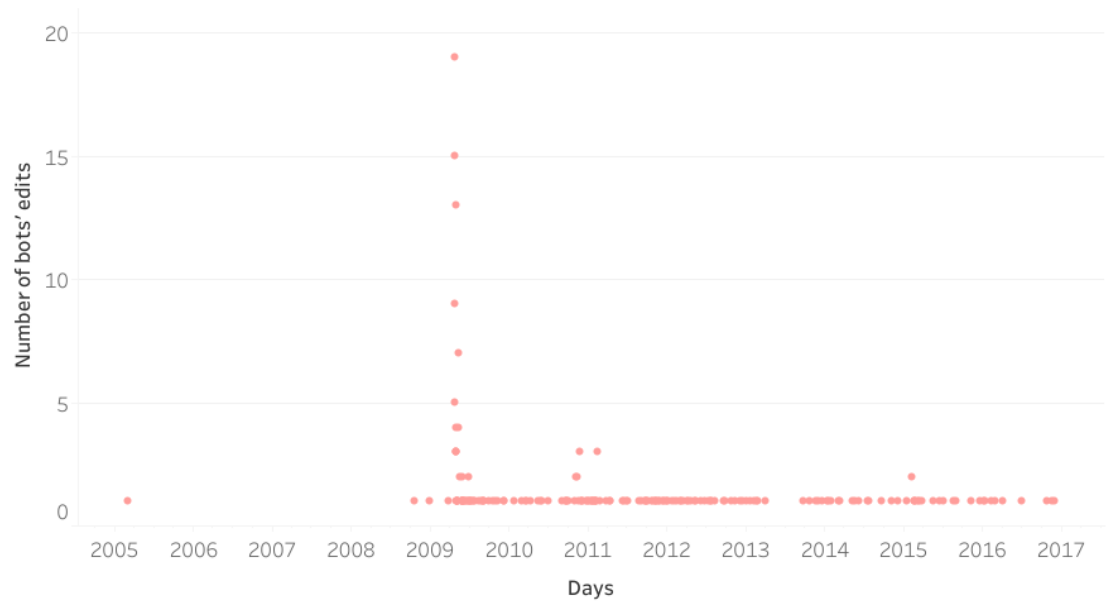
FIGURE 7.11: Bots' Edits in the Swine Influenza Article
https://public.tableau.com/views/Botsswineinfluenza/Sheet2?:embed=y&:
display_count=yes

7.5 Bots' Role During New Disease Outbreaks

This section looks at the role of bots in editing Wikipedia articles during new disease outbreaks. Table 7.5 reveals all the bots that participated in editing the five main articles related to new disease outbreaks (Zika virus, Ebola virus disease, Middle East respiratory syndrome-related coronavirus, Influenza A virus subtype H7N9 and Swine influenza). Various roles and tasks have been undertaken by different bots. These tasks include detecting and reverting vandalism (i.e. by ClueBot NG, DASHBotAV, XLinkBot, VoABot, AntiVandalBot), notifying users about vandalism (i.e. by Cyberbot II), fixing images (i.e. by ImageRemovalBot), fixing templates (i.e. by SporkBot, Robbot), and fixing citations (i.e. by DumbBOT). These roles also include adding and changing interwiki links[3] (i.e. SieBot, VVVBot) as well as maintaing links between different language editions (i.e. YurikBot).

Code	Description
ClueBot NG	Detect and revert vandalism using machine learning algorithms
CitationCleanerBot	Clean citation templates
AnomieBOT	Correct reference syntax errors and attempt to recover orphaned refs from page history
Yobot	Replace and link unlinked fictional characters and missing articles' categories
Citation bot	Replace inconsistent templates such as when pages use 'citation' and 'cite journal' templates (which produce different output styles)
MusikBot	Remove protection templates from unprotected page
BattyBot	Use AutoWikiBrowser auto tagging and general fix functionality
ImageRemovalBot	Periodically go through the deletion log, look for images that have been deleted but not removed from articles and remove them
BG19bot	Add the listas parameter to the talk page of the biography articles that lack it
SporkBot	Orphan or substitute templates
Dexbot	Link to relevant discussions

Continued on next page

---

[3]https://en.wikipedia.org/wiki/Help:Interwiki_linking

Bot	Function
Addbot	Create links between items that it finds while trawling wikipedia articles
Luckas-bot	Add, modify and remove interwikis
Bender the Bot	Convert HTTP → HTTPS for Internet Archive links
Hazard-Bot	Add categories to pages
Monkbot	Cite templates with deprecated parameters
DumbBOT	Remove protected and similar templates from pages that are no longer protected
Bibcode Bot	Find bibcodes and dois for citations
FrescoBot	Run simple tasks like fixing wikilink syntax and satisfing requests inserted in Commons:Bots/Work requests
Legobot	Substitute templates
Snotbot	Search for newly created articles which have fallen off the backlog of Special:Newpages (i.e. they have existed for more than 30 days and were not patrolled by anyone)
JAnDbo	Do utonomously interwiki through all pages (starting from cs, csb, hsb, dsb and others) and do interwiki solving (removing) with assistance of owner
Helpful Pixie Bot	Make repetitive edits that would be extremely tedious to do manually in general such as fixing links and typographical errors
RjwilmsiBot	Set page ranges within page parameter of citation templates to use en-dashes
EmausBot	Interwiki search using different methods
SmackBot	Replace unreferenced tags
CarsracBot	Add and fix interwiki links
DASHBotAV	Catch and revert cases of obvious vandalism
BenzolBot	Help to make many small changes that would take a long time for a person to do alone in general
SieBot	Add and fix interwiki links

Continued from previous page

Bot	Function
XLinkBot	Revert spam and addition of external links which probably do not comply with policy and/or guideline
Dinamik-bot	Add and edit interwiki links
DrilBot	Clean up various common errors in articles using the lists at WikiProject Check Wikipedia
SoxBot	Convert HTML tags to Wiki syntax, fix links to disambiguation pages, find external links that are broken, add categories to pages, put orphan template onto all pages that have no orphan template on it already
VoABot	Watch certain pages and revert edits using regexps denoted as blacklisted
Robbot	Automate template addition, particularly WikiProject templates to talk pages
Obersachsebot	Mostly add and modify interwiki links
PDFbot	Update filesize for external links tagged with PDFlink
Lightbot	Perform janitorial edits related to units and dates
Thijs!bot	Add interwiki links
DOI bot	Add DOIs to citations provided using cite journal
Chobot	Fix interlanguage links
DumZiBoT	Fix incorrect Ref Syntax
PbBot	Perform any merging/moving/emptying required for Wikipedia categories for discussion
CounterVandalismBot	Revert vandalism
RussBot	Move pages from redirected categories to correct categories
MalafayaBot	Perform automatic interwiki exchange
AntiSpamBot	Revert spam links inserted into Wikipedia
demonBot2	Replace old templates with their newer replacements
STBotD	Search and fix missing or changed interwiki links

Bot	Function
Martin's bot	Replace the values of a parameter to allow the external links to continue to work as well as search and fix missing or changed interwiki links
VVVBot	Put interwiki links on pages
AntiVandalBot	Revert vandalism
STBot	Change or remove categories from pages
Tawkerbot	Add subst on Wikipedia pages
CmdrObot	Make repetitive automated or semi-automated edits that would be extremely tedious to do manually in general such as correcting spellings
YurikBot	Help bring different language Wikipedias together, maintain interwiki links between articles, ensure that all relevant pages in all languages have links to each other
Ligulembot	Maintain templates
Guanabot	Update disambiguation pages
TedderBot	Request UserContribs through API, score and find intersection between two users. Post to user talk page
YFdyh-bot	Maintain interwiki link
ArmbrustBot	Help to clean out Wikipedia:Categories for discussion/Working
Cyberbot II	Notify subscribed users in this list to vandalism levels when a set threshold is reached
AvicBot	Remove test edits
LaaknorBot	Add and fix interwiki links
RscprinterBot	Revert test edits
Sahimrobot	Create interwiki links
TjBot	Add and fix interwiki links
VolkovBot	Add and fix interwiki links
H3llBot	Add categories to the relevant article list contained in the template per request

Continued from previous page

Bot	Function
TXiKiBoT	Add and correct interwiki links
D'ohBot	Modify interwiki links
Xqbot	Put db-talk onto orphaned talk pages
FoxBot	Add and fix interwiki links
GrouchoBot	Add and fix interwiki links
LucienBOT	Add and fix interwiki links
SilvonenBot	Add and fix interwiki links
DSisyphBot	Add and fix interwiki links
GhalyBot	Add interwiki links between articles
Rubinbot	Add interwiki links
Cydebot	Make repetitive automated or semi-automated edits that would be extremely tedious to do manually in general
Numbo3-bot	Add and fix interwiki links and update templates
Synthebot	Add and fix interwiki links
Alexbot	Add and fix interwiki links
ArthurBot	Add and fix interwiki links and link templates
Broadbot	Update and create interwikilinks
BepBot	Add and fix interwiki links
Idioma-bot	Add and fix interwiki links
BotKung	Add interwikies between different languages of Wikipedia pages
Zorrobot	Add and fix interwiki links
AkhtaBot	Add and fix interwiki links
Almabot	Add and fix interwiki links
Muro Bot	Fix double redirects

Table 7.5: Bots' Role During New Disease Outbreaks

7.6   Summary

This chapter suggested that articles related to Zika, Ebola, and swine influenza have experienced high to moderate burstiness, while articles related to MERS and H7N9 have experienced low burstiness. As such, this chapter aimed to explore factors that might influence patterns of editing behaviours on Wikipedia during new disease outbreak crises (RQ3). This difference in the level of burstiness could be explained by many factors such as the amount of press coverage and the number of WHO warnings that some diseases are public health emergencies of international concern. In other words, Wikipedia articles related to Zika, Ebola, and swine influenza experienced a higher level of burstiness because these new disease outbreaks were reported widely in the press and were declared as public health emergencies. This chapter also explored whether the surge in the number of bots' edits might be another factor that could influence the level of burstiness in articles related to new disease outbreaks. However, even through that bots have undertaken various repetitive automated or semi-automated edits that would be extremely tedious to do manually, the data showed that bots play a trivial or no role in triggering burstiness during new disease outbreak crises. After identifying and explaining factors that influence attention to Wikipedia articles during crises, the following chapter aims to look at other forms of responses to crises on Wikipedia. This is done by discussing forms and strategies for managing the uncertainty during new disease outbreak crises.

Chapter 8

Results: Response to Uncertainty in Wikipedia

'Medicine is a science of uncertainty and an art of probability.'

— Sir William Osler
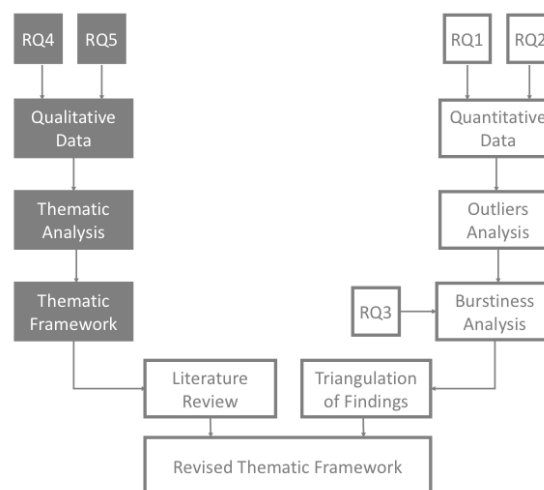
FIGURE 8.1: Chapter's Placement within Wider Research Design

The previous two chapters aimed at understanding the Wikipedia editors' response to crises by looking at the patterns of editing behaviours on articles related to new disease outbreaks. This was achieved by identifying the deviation in the editing activities during new disease outbreaks as well as by finding factors that could explain this deviation. This chapter, however, aims to focus on understanding the Wikipedia editors' response to the uncertainty during crises. This was achieved by looking at the types of uncertainty as expressed by Wikipedia editors (RQ4) and the strategies used to manage different forms of uncertainty (RQ5). Similar to the previous two chapters, this chapter compares the

forms and the strategies to manage the uncertainty in articles related to new disease outbreaks with the forms and the strategies to manage the uncertainty in articles related to chronic diseases. The chapter concludes by presenting a framework that consolidates the forms and the strategies to manage the uncertainty during crises.

8.1   Wikipedia Talk Pages

The methods chapter (Chapter 5) illustrated that qualitative data sources were acquired from Wikipedia talk page achieves. Table 8.1 reports the number of posts acquired from each of the 10 main articles included in the sample. In all, 1686 posts are further assessed and examined throughout this chapter.

Articles	Number of Posts
Zika Virus	56
Ebola Virus Disease	247
MERS	21
H7N9	23
Swine Influenza	130
Diabetes	291
Obesity	370
Cancer	344
Asthma	171
Cardiovascular Disease	33
New Disease Outbreaks	477
Chronic Diseases	1209
Total Number of Posts	1686

TABLE 8.1: Total Number of Posts in Each Wikipedia Article

Article	Edit History	Talk Pages	Talk Pages % of Edit History
Zika virus	559	76	14%
Ebola	3566	415	12%
MERS	351	28	8%
H7N9	219	31	14%
Swine influenza	1589	433	27%

TABLE 8.2: Proportion of Editors Participating on Talk Pages

Table 8.2 shows the proportion of editors who post on Wikipedia talk pages in relation to the proportion of editors who contribute to each of the main articles. These data unveil that the number of editors who post on Wikipedia talk pages is very small in comparison to the number of editors who contribute to Wikipedia articles. For example, the number of editors who post on the Zika virus talk page represents only 14% of the

number of editors who contribute to the Zika virus article.

The following reports the qualitative results obtained using the thematic analysis described in Chapter 5.

8.2 Types of Uncertainty

It is unquestionable that editors experience and express different forms of uncertainty while trying to edit Wikipedia articles. Uncertainty is not exclusively perceived when editing articles related to new disease outbreaks; it is also perceived when working on articles related to chronic diseases. This section aims to define the uncertainty in Wikipedia. This was done by identifying the different forms of uncertainty as reflected from the editors' discussions in the talk pages. The forms of uncertainty were grouped under three overarching themes that emerged from an analysis of the talk pages of each of the main articles related to new disease outbreaks and chronic diseases. These themes were scientific uncertainty, references uncertainty, and conflicting references uncertainty.

8.2.1 Scientific Uncertainty

Wikipedia editors notice that some information about prevention, causes, symptoms, and treatments of diseases are uncertain. This uncertainty is perceived when it is difficult to obtain scientific facts, particularly when scientific knowledge is still in the early stage of research and development. As a result, this uncertainty is expressed in Wikipedia when scientific information is reported in the references as uncertain or if the relevant references are either absent or poor. As reported in Table 8.3, scientific uncertainty is the most common type of uncertainty that emerges in the talk pages of articles related to both new disease outbreaks and chronic diseases (69%). The following are the main examples of scientific uncertainty as retrieved from the talk pages of articles related to new disease outbreaks and chronic diseases. These examples are accompanied with reasons that justify the classification of these examples under scientific uncertainty.

8.2.1.1 New Disease Outbreaks

Zika Virus:
During the Zika virus outbreak, one of the Wikipedia editors posted a request to update the information about *Aedes albopictus*. This was because it was assumed by the editors that this type of mosquito could spread the Zika virus. However, another editor replied that:

CometEncke thanks for the pointer. I was the one that added the (!) primary studies = not considered WP: MEDRS :-) on Aedes albopictus and Zika, because I find it a compelling piece that hasnt been carved out by popular media. there isnt much evidence (.yet)–5 February 2016 (UTC)

Classification Reason: Relevant references are either absent or poor.

The other editor expressed doubts about the transmission of the Zika virus from the mother to the baby:

While the virus "may" spread from the mother to the baby during pregnancy I am not seeing evidence that this always occurs or that if it occurs it always results in severe problems–15 April 2016 (UTC)

Classification Reason: Relevant references are either absent or poor.

Different editors noted the uncertainty over the link between the Zika virus and microcephaly, which is a birth defect that occurs when a baby's head is smaller than the heads of other babies of the same sex and age[1]

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–30 January 2016 (UTC)

Classification Reason: Scientific information is reported in the references as uncertain.

Ebola Virus Disease:
Different forms of disease uncertainty emerged during the 2014 Ebola Virus Disease outbreak. Some of these forms can be recognised from the talk pages of the Ebola Virus Disease articles. For instance, a group of editors were uncertain whether specific species of fruit bats are a natural host of Ebola:

There seems to be growing evidence for several species of fruit bats to be a natural reservoir of Ebola, though it is unclear if othere species are involved in the transmission of Ebola to primates–11 November 2008 (UTC)

---

[1]https://www.cdc.gov/ncbddd/birthdefects/microcephaly.html

Classification Reason: Scientific information is reported in the references as uncertain.

Different posts showed that the editors had expressed uncertainty about the lengths of the Ebola virus incubation period:

> *With respect to the question of the gold standard of 21 days, there has been some recent controversy about that and some have called for longer incubation periods since a small proportion of cases do seem to occur past 21 days–9 November 2014 (UTC)*

Classification Reason: Scientific information is reported in the references as uncertain.

Other editors perceived the information about the immunity of the Ebola virus to be uncertain:

> *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 considerations–7 October 2014 (UTC)*

Classification Reason: Relevant references are either absent or poor.

Replying to the previous post, an editor asserted that:

> *I took the liberty of placing your question in the correct order. As for the question, I'm doubtful that any significant study has been made, all medical personnel are still involved in treating, triaging, supporting and generally scrambling to try to get control of the situation. Statistical analysis is for after the epidemic is brought under control–8 October 2014 (UTC)*

Classification Reason: Relevant references are either absent or poor.

H7N9:
The variations of uncertainty became visible on Wikipedia during the 2013 H7N9 disease. For example, the uncertainty about the age of people affected by the virus was expressed on Wikipedia:

*While it is, at least theoretically, possible that males are biologically vulner-
able to H7N9, you are correct that no experts have made that claim. Re-
searchers have suggested a number of theories, such as that elderly Chinese
males may have more exposure to poultry or birds than other groups. One
widely commented upon theory is in (Skowronski 2013). The last paragraph
of the Transmission section in the main article discusses that theory a bit,
using a decent reference–3 May 2013 (UTC)*

Classification Reason: Scientific information is reported in the references as
uncertain.

Swine Influenza:
Wikipedia editors actively posted on the Wikipedia talk pages to report and discuss
the different forms of scientific uncertainty during the swine influenza outbreak. For
example, an editor was not sure whether the human-transmissible strain of the virus
should be called 'swine influenza':

*Hey there. Yes, I do see your argument, but unfortunately with organisations
like the CDC and WHO all referring to this as "swine inuenza" (see WHO for
example (http://www.who.int/csr/disease/swin eu/en/index.html)) we have
to do the same on Wikipedia. Once we have more information on this H1N1
strain we could split the articles into one on this human-transmissible strain
and a second more general one about "swine inuenza", but presently we're
stuck with this nomenclature–29 April 2009 (UTC)*

Classification Reason: Relevant references are either absent or poor.

Another editor reported the uncertainty over the cases and the death estimates of swine
influenza:

*Accurate estimates for cases and deaths from this pandemic (from swine-
origin pandemic H1N1/09 [the 1918 pandemic was also H1N1, so the "09"
is needed to distinguish it]) do not yet exist. By sometime next year we will
have numbers that can be compared to existing accurate estimates for seasonal
flu. The lab confirmed cases and deaths are only useful for tracking spread
and mutations, not numbers of cases and deaths–2 November 2009 (UTC)*

Classification Reason: Relevant references are either absent or poor.

Moreover, one of the editors commented that it is still scientifically unknown whether
certain vaccines for swine influenza are effective to contain the spread of the virus:

> *For now, whether it will be contained, and if so exactly when may become more contagious and break out, when it will get Tamiflu-resistant, and when a vaccine will ship in quantity are known unknowns, and evolving developments to watch–1 May 2009 (UTC)*

Classification Reason: Relevant references are either absent or poor.

8.2.1.2 Chronic Diseases

Diabetes:

Research around the chronic conditions of Type 1 and Type 2 diabetes is still ongoing. Therefore, scientific information related to diabetes remains uncertain. Only a few of these uncertainties are expressed on Wikipedia. For example, one of the editors expressed uncertainty over the cause of diabetes mellitus Type 2:

> *This kind of victim blaming isn't what people need. "Diabetes mellitus type 2 is presently of unknown etiology" which means that medical science has yet to figure out what causes it. Without some pretty stunning verification, statements like "…obesity is not the only cause of Type II but it is the leading cause" will not add anything of value to the article–5 February 2007 (UTC)*

Classification Reason: Relevant references are either absent or poor.

Moreover, one editor justified his answers in another post by stating the scientific uncertainty about the relationship between testosterone and diabetes:

> *Cortisol is consensually recognized as a glucose regulator and there is literature debating its relationship with diabetes, see also Cushing syndrome, even if the picture is very complex and it is not easy to figure out what is the state of the consensus. On testosterone, the picture from the links above looks much more fuzzy; there is consensus on the existence of a relationship between testosterone levels and diabetes, but I can't find consensus about a causative relationship, and only an abstract about the possibility of testosterone therapy. There is no evidence to conclude that testosterone/cortisol are consensually considered the main causative agents of diabetes, nor that subclinical Cushing is consensually considered a leading cause of it–25 October 2009 (UTC)*

Classification Reason: Relevant references are either absent or poor.

Another editor clarified the uncertain issues related to pancreas transplants as a cure for diabetes:

I can only speculate as to the reasoning, but I presume the problem lies with the difficulty of getting transplants and the complications that can ensue afterwards. I don't think anyone would claim that we have a cure for heart disease because we can perform heart transplants. Additionally, I believe there is little understanding as to what may trigger another beta cell attack - for type 1 - and the risk of subsequent beta cell failure as a result of insulin resistance-induced stress - for type 2 - is not removed–9 November 2007 (UTC)

Classification Reason: Relevant references are either absent or poor.

Furthermore, a different editor expressed that it is difficult to confirm that there is a relationship between diabetes and pancreatic cancer:

I'm just a layman with an interest in biosciences, but diabetes has been one area I happened to know a lot more since many of my relatives seem to have it and so I stumbled upon pancreatic cancer patients too. Some doctors here swear that one is doomed to get diabetes once pancreatic cancer "sets in" or "attacks" (since I don't know the right medical term). Others are not convinced–Jun 24, 2005 (UTC)

Classification Reason: Scientific information is reported in the references as uncertain.

Likewise, other editors expressed doubt over calling the gastric bypass surgery a 'cure':

I think this is too stark an answer. Various surgeons in various places have patients who seem to have been cured. And further seem to have been so as a result of a particular sort of bariatric surgery. However, not all such surgeries are done the same way (even the cited type), and not all hospital courses are identical, and since we don't really understand just what gives with the insulin resistance of assorted Type 2 (if it is the same from one to the next, in any case and if the effect is lasting), it is too much to claim in a WP article that there is a cure. This word means much to the Average Reader and something rather different to the readers of such as NEJM–15 October 2007 (UTC)

Classification Reason: Scientific information is reported in the references as uncertain.

Obesity:

Similar to diabetes, obesity healthcare providers are frequently placed in situations where a well-defined answer does not exist. This can be seen on the Wikipedia talk pages where editors discuss different obesity-related uncertainties. Examples include the uncertainty over the role of genetics in causing obesity:

> *Quite evidently individuals with the same level of physical exercise and same intake of food may have remarkably different BMI. Argue that this may be due to lean body mass/total body mass ratio contributing to different basal metabolism levels etc., but we cannot be sure, and even then LBM is just one of many factors where genetics play a role, and to dismiss genetics is not viable–31 May 2014 (UTC)*

Classification Reason: Relevant references are either absent or poor.

A different post included a question about the role of insulin resistance in causing obesity:

> *The state of overfatness/obesity clearly contributes to insulin resistance, which in turn can cause type 2 diabetes. Virtually all obese and most type 2 diabetic individuals have marked insulin resistance. Although the association between overfatness and insulin resistance is clear, the exact (likely multifarious) causes of insulin resistance remain less clear. Importantly, it has been demonstrated that appropriate exercise, more regular food intake and reducing glycemic load all can reverse insulin resistance in overfat individuals (and thereby lower blood sugar levels in those who have type 2 diabetes). Again, this seems to pertain to a therapeutic intervention (obese people have insulin resistance, and a diet helps improving this). Or is the intent to show that insulin resistance causes obesity?? This has been suggested in the literature but is by no means an accepted theory–6 November 2005 (UTC)*

Classification Reason: Scientific information is reported in the references as uncertain.

The same post also included a question about the role of the leptin hormone in causing obesity:

> *Overfatness can unfavourably alter the hormonal and metabolic status via resistance to the hormone leptin, and a vicious cycle may occur in which insulin/leptin resistance and overfatness aggravate one another. The vicious cycle is putatively fuelled by continuously high insulin/leptin stimulation and fat storage, as a result of the high intake of strongly insulin/leptin-stimulating*

foods and energy. Both insulin and leptin normally function as satiety signals to the hypothalamus in the brain; however, insulin/leptin resistance may reduce this signal and therefore allow continued overfeeding despite large body-fat stores. In addition, reduced leptin signalling to the brain may reduce leptin's normal effect to maintain an appropriately high metabolic rate. Leptin is actually discussed further down in the article, and whether the resistance to it is a cause is still a matter of debate. The role of insulin as an appetite regulator is similarly debated and cannot be presented as fact–6 November 2005 (UTC)

Classification Reason: Scientific information is reported in the references as uncertain.

Cancer:

Cancer-related uncertainties were also manifested on the Wikipedia talk pages. These included noting the limited scientific evidence of the role of certain diets in causing cancer:

Currently we state "While many dietary recommendations have been proposed to reduce the risk of cancer, they are based on relatively limited evidence." This is supported by this 2011 review which states "This stands in sharp contrast to the paucity of evidence, most of which is derived from epidemiological data. Controlled trials, on the other hand, are mostly missing. However, there is no convincing evidence that fat uptake per se (i.e. independent from total energy uptake and obesity) is a risk factor for malignant disease. There is currently no compelling evidence that supplementing vitamins, antioxidants or other micronutrients reduces cancer incidence"–25 July 2013 (UTC)

Classification Reason: Relevant references are either absent or poor.

Additionally, another editor found information about the role that the DNA repair plays in cancer to be uncertain:

These are controversies because the precise details are not understood. To see what I mean, contrast this with something like the citric acid cycle, where the precise details are all worked out. In your examples, DNA repair, apoptosis, etc have been definitively established, but the role that they play in cancer is not well established–23 Jun 2005 (UTC)

Classification Reason: Scientific information is reported in the references as uncertain.

The talk pages of the cancer article also included doubts about the role that Vitamin D and sun exposure play in cancer:

> *There is a correlation between increased sun exposure and a 30-50% risk reduction. Does going out in the sun cause the reduction, or are there confounders (the most obvious of which is physical activity, which clearly reduces cancer risk and is conceivably associated with greater sun exposure)? There's certainly suggestive evidence that vitamin D (and, indirectly, sun exposure) are responsible, and some studies have tried to control for activity level, etc., so I'm not blowing that off. But the transition from "Sun exposure is correlated with a risk reduction" to "sunlight prevents cancer" is what I'm wary of. We may be heading in that direction, and there's certainly biological plausibility to the idea that vitamin D has an anticancer effect, but I don't think we're all the way there yet–2 March 2007 (UTC)*

Classification Reason: Relevant references are either absent or poor.

Asthma:

Particular details about the disease asthma can be ambiguous, complex, unpredictable, or probabilistic. This can lead to the ambiguity and uncertainty becoming evident in the talk pages of the asthma Wikipedia article. For instance, a group of editors were uncertain about whether pollution can be considered a cause of asthma:

> *Research on the degree to which industrial pollution can be a principle cause of asthma is ongoing, but existing data strongly suggest that industrial pollution is among the environmental factors that can trigger asthmatic episode–20 January 2006 (UTC)*

Classification Reason: Scientific information is reported in references as uncertain.

Similarly, an editor found the role that the exposure to endotoxin plays in asthma to be uncertain:

> *Liu's paper (https://www.ncbi.nlm.nih.gov/pubmed/....) is a little old now (2004). The role of endotoxin in asthma remains unclear, perhaps even controversial. This 2011 paper (https://www.ncbi.nlm.nih.gov/pmc/....) describes the current understanding, which hasn't changed much–26 November 2013 (UTC)*

Classification Reason: Scientific information is reported in the references as uncertain.

8.2.2   References Uncertainty

Finding references to verify claims on Wikipedia articles is considered one of the main responsibilities of Wikipedia editors. However, interpreting the information from different references is considered a challenging task. Thus, references uncertainty is defined as the perceived ambiguity about information or claims mentioned in particular references. This pushes editors to post on Wikipedia talk pages in order to clarify the claims within the references, to check whether certain references are valid, or to simply question the claims from some references. While scientific uncertainty refers to the ambiguity about scientific information in general, references uncertainty focuses on the ambiguity present in specific references and publications. References uncertainty could involve the editors' assessment of the conclusions presented in the publications as well as an assessment of the study design and sample size, and that of whether there are similar studies that agree with the given conclusions. As illustrated in Table 8.3, examples of references uncertainty can be observed in the talk pages of the articles related to new disease outbreaks (10%) and in those of the articles related to chronic diseases (15%). The following section reports the examples of references uncertainty with reasons that justify classifying these examples as such.

8.2.2.1   New Disease Outbreaks:

Ebola Virus Disease:
The Ebola Virus Disease article is the only article on new disease outbreaks that includes different forms of references uncertainty. The following quote illustrates that Wikipedia editors found historical information about a previous Ebola outbreak mentioned in a paper to be uncertain:

> *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–October 2014 (UTC)*

Classification Reason: Questions claims from some reference.

Another editor questioned a claim about the Ebola virus incubation period that originated from the World Health Organisation (WHO) and Centers for Disease Control and Prevention (CDC):

> *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–8 November 2014 (UTC)*

Classification Reason: Questions claims from some reference.

The previous quote suggests that some uncertainties were perceived on references that belong to authoritative sources such as WHO and CDC.

8.2.2.2 Chronic Diseases:

Diabetes:
References uncertainty emerged as one of the types of uncertainty that was expressed in the talk pages of the articles related to chronic diseases. This included the talk pages of the diabetes article as one of the editors had questions about the information related to the possible causes of diabetes that were cited in an academic paper:

> *Even in people with a normal range fasting blood sugar, there is a direct relationship with the risk of diabetes. doi:10.1016/j.amjmed.2008.02.026 Does this then mean (or hint) that glucose itself causes tissue changes which lead to the Type 2. For instance, changes in the upper small intestine which then proceed to emit a signal which causes the insulin resistance in the periphery and thus the cascade of events which is too often Type 2?–21 May 2008 (UTC)*

Classification Reason: Clarify claims within references.

Another editor questioned the information published in The New England Journal of Medicine. This led to a long discussion to clarify and discuss whether surgery could cure Type 2 diabetes:

> *Improvement achieved in obese patients through bypass surgery does not necessarily imply that same can be achieved in non-obese patients. When weighing a claim from a study, several factors need to be considered other than those already discussed; such as what is the size of the study sample, is it randomised, is it blinded, is it a prospective study. Let me roughly calculate minimum number of subjects required to assess a new treatment for type 2 diabetes using this simple (http://www.surveysystem.com/....) formula. DM2 has more than 100 million sufferers worldwide, as such minimum 384 subjects are required to be 95% confident that error margin would be less than 5% (standard confidence interval). Some large scale trials such as NEJM*

one did show long term good blood sugar control after bariatric surgery. But like NEJM study, they are done on obese patients. GI bypass surgery is an accepted form of treatment in patients whose BMI is more than 35. I am not aware of any large scale randomised study on non-obese patients–28 September 2007 (UTC)

Classification Reason: Clarify claims within references.

Obesity:

Specific publications about the medical condition of obesity were considered dubious by Wikipedia editors. Therefore, different forms of references uncertainty can be recognised from the discussions on the obesity article talk pages. For instance, some editors were uncertain about the relationship between obesity and some viruses, as suggested in one of the academic studies:

The only viruses linked serologically to human obesity are an avian virus (SMAM-1,PMID 9385623 (https://www.ncbi.nlm.nih.gov/....)) and AD-36 (Int. J. Obesity 22:S57 and FASEB J 3:A230). The studies are conjecturous, have small numbers, and oddly all originate from the same research group in Wisconsin–7 November 2005 (UTC)

Classification Reason: Questions claims from some references.

In a different post, an editor reported a study to suggest that there is a relationship between physical activity and obesity. However, other editors were uncertain about this relationship as similar studies need to be conducted and reproduced:

"Fatness leads to inactivity, but inactivity does not lead to fatness: a longitudinal study in children (EarlyBird 45)" http://adc.bmj.com/content/early/.... "This reverse causality may explain why attempts to tackle childhood obesity by promoting PA have been largely unsuccessful." — A single study. Now let's wait for this to be reproduced–10 July 2010 (UTC)

Classification Reason: Questions claims from some references.

Cancer:

Editors were also skeptical about certain studies published in academic papers such as studies that focused on the link between obesity and different types of cancers:

There is a well-established link between obesity and certain cancers. Excess carbohydrate intake is linked with obesity, as well with as lower socioeconomic groupings which have other cancer epidemiological associations, such

> *as increased smoking and alcohol use, industrial exposure etc etc. These are*
> *all quite difficult to disentangle, and I am sceptical about the Mexican paper,*
> *which is quite underpowered from an epidemiological p.o.v. and which has*
> *a debateable study design. The Korean paper is interesting, and shows an*
> *association, but not necessarily a causal link–19 February 2008 (UTC)*

Classification Reason: Questions claims from some references.

In addition, the editors were skeptical about a particular research paper that included questionable materials about the causes of cancer:

> *The investigative work in this paper may be sound, but the conclusion is*
> *fairly speculative (this is clear when reading the peer-reviewed paper which*
> *uses much more nuanced terms than the accompanying press release) because*
> *most of the sample were relatively young and not all cancers will be evident*
> *in remains (in bone samples, for example, they were mainly hoping to find*
> *osteosarcomas, which are rare)–30 November 2010 (UTC)*

Classification Reason: Questions claims from some references.

Asthma:

Editors were uncertain about some information found in the publications related to asthma. For example, editors expressed references uncertainty when coming across a study that suggests that there is a link between pollution and asthma:

> *It only demonstrates a strong correlation, rather than provides proof of pri-*
> *mary causation (but that is not to say a correlation is not important to take*
> *note of). Even if the effect should be proven, this study does not clarify*
> *whether pollution increases the de-novo incidence of asthma, or raises the*
> *profile of asthma by exacerbating established cases (either of otherwise mild*
> *sub-clinical asthma to a state that it gets diagnosed, or of exacerbating es-*
> *tablished cases causing greater usage of medical resources)–17 January 2006*
> *(UTC)*

Classification Reason: Questions claims from some references.

8.2.3 Conflicting References Uncertainty

As discussed previously, editors spend time and effort in trying to evaluate the information from different references or publications. In some cases, the editors find contradictory information about the diseases cited in the different references. This is described as

conflicting references uncertainty, and this form of uncertainty motivates the editors to either seek clarification or defend their claims on the Wikipedia talk pages. In contrast to references uncertainty, which covers the uncertainty about the information cited in specific publications, conflicting references uncertainty focuses on the conflicting and incompatible claims mentioned in multiple publications as well as in the literature in general. The analysis reveals that cases of conflicting references uncertainty are found in the talk pages of articles related to new disease outbreaks and chronic diseases. However, Table 8.3 shows that instances of conflicting references uncertainty are observed more frequently in the talk pages of the articles related to new disease outbreaks (28%) than in the talk pages of the articles related to chronic diseases (12%). The following section delineates the instances of conflicting references uncertainty followed by the reasons to classify these instances as such.

8.2.3.1   New Disease Outbreaks

Ebola virus disease:
Noticeably, editors find and discuss different cases of conflicting references in the talk pages of the Ebola virus disease articles. This includes finding publications that report different start dates of the Ebola virus disease outbreak:

> *For the index case in Ebola virus disease2013_to_2014_West_African_outbreak, we use the date 28 December 2013 which is supported by the WHO source [23] (http://who.int/csr/....) but the other source uses 6 December 2013.[24] (http://www.nytimes.com/2014/08/10/world/....) In Ebola virus epidemic in West AfricaOutbreak and 2014 Ebola virus disease epidemic timeline December 2013 we use the 6 December 2013 date per the sources there. Which is it, or do we quote both dates?–31 October 2014 (UTC)*

> **Classification Reason**: Conflicting and incompatible claims mentioned in publications.

Cases of conflicting references uncertainty also include finding publications that report contradictory information about the transmission of the Ebola virus:

> *The citation links to a World Health Organization page with one bullet point stating that transmission without direct contact hasn't been documented an that airborne transmission "..has not been documented during previous EVD outbreaks".[2] However in November of 2012, Scientific Reports, a research outlet from the publishers of Nature, published a paper about Ebola transmission from pigs to non-human primates. It included reason to believe that Ebola can travel by air and not only by close contact or body fluid exchange–6 August 2014 (UTC)*

Classification Reason: Conflicting and incompatible claims mentioned in publications.

Additionally, cases of conflicting references uncertainty include expressing doubts about contradictory information cited by authoritative sources:

> *The CDC briefing states that "Individuals who are not symptomatic are not contagious." However, this is simply wrong. Symptoms facilitate easier transmission, but according to the WHO "People are infectious as long as their blood and secretions contain the virus" (http://www.who.int/mediacentre/....)– 22 August 2014 (UTC)*

Classification Reason: Conflicting and incompatible claims mentioned in publications.

Moreover, conflicting references uncertainty entail the editors discussing contradictory information about the number of cases affected by the Ebola virus disease outbreak:

> *I have revised the number of cases for the Guinea and Sierra Leone outbreak shown in the table. The source cited, Fox News (reliable?) mentions 78 deaths from 122 cases of suspected Ebola since January, up from 70. Of these, there were 22 laboratory confirmed cases. The World Health organization states, 'As of 28 March 2014 (there are), 103 suspected and confirmed cases, including 66 deaths, (CFR 64%)'–31 March 2014 (UTC)*

Classification Reason: Conflicting and incompatible claims mentioned in publications.

H7N9:
Similar to the Ebola virus disease article, Wikipedia editors questioned the conflicting information about the number of cases affected by 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–6 May 2013 (UTC)*

Classification Reason: Conflicting and incompatible claims mentioned in publications.

Swine influenza:

It is not always easy to find consistency among different reporting agencies as some of the inconsistency can be justified by the timing and the release of the 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...–4 May 2009 (UTC)*

Classification Reason: Conflicting and incompatible claims mentioned in publications.

This example acknowledges the constantly changing nature of scientific information, particularly during crises and new disease outbreaks.

8.2.3.2 Chronic Diseases

Diabetes:

The Wikipedia talk pages of chronic diseases including diabetes contain several cases of conflicting references uncertainty. The following post illustrates that an uncertainty could exist around references that debate whether there is a relationship between cortisol and diabetes:

> *Cortisol is consensually recognized as a glucose regulator and there is literature debating its relationship with diabetes, see also Cushing syndrome, even if the picture is very complex and it is not easy to figure out what is the state of the consensus–25 October 2009 (UTC)*

Classification Reason: Conflicting and incompatible claims mentioned in publications.

A similar post revealed that references uncertainty can be observed on Wikipedia as studies could cite contradictory information about the correlations between diabetes and breastfeeding:

> *Some articles say there is 2% reduction in risk of type 1 if you breastfeed. Other articles say that it is inconclusive–10 November 2006 (UTC)*

Classification Reason: Conflicting and incompatible claims mentioned in publications.

Obesity:

The talk pages of the obesity Wikipedia article also contain the evidence of conflicting references uncertainty. For instance, one of the editors declared that there is no consensus on the relationship between weight and food type consumption:

> *The evidence is contradictory when it comes to the effect of food type consumption on weight lose / gain–21 September 2008 (UTC)*

Classification Reason: Conflicting and incompatible claims mentioned in publications.

Another editor emphasised that there is a disagreement between different studies that investigate the relationship between obesity and physical activity:

> *Other studies disagree...–11 July 2010 (UTC)*

Classification Reason: Conflicting and incompatible claims mentioned in publications.

Asthma:

Wikipedia editors also pinpoint the contradictory information cited in articles about asthma. For example, a Wikipedia editor presented some contradictory findings about the correlations between the prevalence of asthma conditions and socioeconomic factors:

> *Hi there! I'm sorry I missed this message earlier. You're absolutely right, that sentence is misleading, and I'd have rewritten it if I'd seen it. There have been varying reports on how socioeconomic status correlates with asthma prevalence: some studies indicate no correlation, others a negative one, yet others a positive one. A problem with the data is that many of the earlier studies done on this issue were of a design that made it easier for spurious associations to be "found". One very large study in Europe recently found that lower socioeconomic status correlates with asthma incidence; however, a new New Zealand study that used what we call a longitudinal design found no real relation between s-e status and the risk of getting asthma, when possible confounding factors had been controlled for. So it's something of an open issue–7 October 2005 (UTC)*

Classification Reason: Conflicting and incompatible claims mentioned in publications.

Cardiovascular Disease:

Similarly, Wikipedia editors discussed some contradictory conclusions about the association between the increasing risk of cardiovascular diseases and the consumption of dietary saturated fat:

> *We have meta analyses that state different things. This one states "that there is no significant evidence for concluding that dietary saturated fat is associated with an increased risk of CHD or CVD"[4] (http://www.ncbi.nlm.nih.gov/....) Doc James (talk · contribs · email) (if I write on your page reply on mine) 01:17, 20 April 2013 (UTC) And than we have this just out " An updated meta-analysis of linoleic acid intervention trials showed no evidence of cardiovascular benefit."[5] (http://www.ncbi .nlm.nih.gov/pubmed/....). Doc James (talk · contribs · email) (if I write on your page reply on mine) 01:26, 20 April 2013 (UTC) Well, the role of saturated fat in CHD is pretty well established (every major medical organization recognizes it as a risk factor for CHD); the more recent research has suggested those who substitute it with carbohydrates, however, do not get much (if any) benefits as compared to replacing it with mono or poly unsaturated fat (see for example these recent reviews [6] (http://www.ncbi.nlm. nih.gov/....) [7] (http://www.ncbi.nlm.nih.gov/....) [8] (http://www.ncbi.nlm.nih.gov/....)). That saturated fat doesn't have any effect on CHD is a WP:FRINGE position, and suggesting there is a signficant controversy here probably gives undue weight to that position–20 April 2013 (UTC)*

Classification Reason: Conflicting and incompatible claims mentioned in publications.
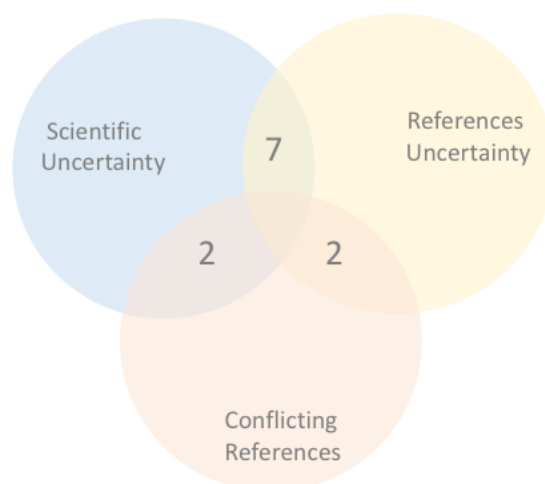
FIGURE 8.2: Overlap Between Different Types of Uncertainty

Articles	Scientific Uncertainty	References Uncertainty	Conflicting References	Uncertainty Relative to Posts
Zika Virus	7	0	0	13%
Ebola Virus Disease	10	4	9	9%
MERS	0	0	0	0
H7N9	3	0	1	17%
Swine Influenza	4	0	1	4%
Diabetes	19	4	2	9%
Obesity	13	4	2	5%
Cancer	16	2	1	5%
Asthma	5	1	3	5%
Cardiovascular Disease	0	0	1	3%
New Disease Outbreaks	24	4	11	8%
Chronic Diseases	53	11	9	6%
Uncertainty	77	15	20	7%

TABLE 8.3: Frequency and Percentage of Each Type of Uncertainty in Wikipedia Articles

Overall, posts that include different types of uncertainty account for only 7% of the total number of posts. Moreover, posts in the talk pages of new disease outbreak articles include a slightly higher percentage of uncertainty relative to the total number of posts (8%) than those in the talk pages of chronic diseases articles (6%). Figure 8.2 confirms that it is possible to observe posts that contain two different types of uncertainty such as posts that contain both scientific uncertainty and references uncertainty. Figure 8.3 to Figure 8.10 exhibit different types of uncertainty over time in articles related to new disease outbreaks and chronic diseases. The length of each timeline reflects findings of different types of uncertainty from 2001 till 2016.

Figure 8.3: Uncertainty over Time in the Zika Article
https://public.tableau.com/views/ZikaArticleUncertainty/Dashboard1?:
embed=y&:display_count=yes&:origin=viz_share_link

Figure 8.4: Uncertainty over Time in the Ebola Virus Article
https://public.tableau.com/views/EbolaWikipediaArticleUncertianty/
Dashboard2?:retry=yes&:embed=y&:display_count=yes&:origin=viz_share_link

Figure 8.5: Uncertainty over Time in the H7N9 Article
https://public.tableau.com/views/H7N9WikipediaArticleUncerainty/
Dashboard1?:embed=y&:display_count=yes&:origin=viz_share_link

Figure 8.6: Uncertainty over Time in the Swine Influenza Article
https://public.tableau.com/views/SwineInfluenzaArticleUncertainty/
Dashboard1?:retry=yes&:embed=y&:display_count=yes&:origin=viz_share_link

FIGURE 8.7: Uncertainty over Time in the Diabetes Article
https://public.tableau.com/views/DiabetesWikipediaArticleUncertainty/
Dashboard1?:retry=yes&:embed=y&:display_count=yes&:origin=viz_share_link

FIGURE 8.8: Uncertainty over Time in the Obesity Article
https://public.tableau.com/views/ObesityWikipediaArticleUncertainty/
Dashboard1?:retry=yes&:embed=y&:display_count=yes&:origin=viz_share_link

FIGURE 8.9: Uncertainty over Time in the Cancer Article
https://public.tableau.com/views/CancerWikipediaArticleUncertainty/
Dashboard1?:retry=yes&:embed=y&:display_count=yes&:origin=viz_share_link

FIGURE 8.10: Uncertainty over Time in the Asthma Article
https://public.tableau.com/views/WikipediaArticles-AsthmaUncertainty/
Dashboard1?:retry=yes&:embed=y&:display_count=yes&:origin=viz_share_link

8.3   Strategies to Manage Uncertainty

As shown in the previous section, Wikipedia editors encounter different sources of uncertainty while editing articles related to new disease outbreaks or articles related to

chronic diseases. As the first step to manage uncertainty, Wikipedia editors resort to the talk pages to ask questions, negotiate, and discuss different sources of uncertainty. After the examination of the posts on the Wikipedia talk pages, the following section outlines multiple strategies used to manage uncertainty. These strategies can be grouped into five categories: rely on authoritative sources, report the uncertainty in the article, ignore the uncertainty, consult experts for advice, and set up a mailing list about ongoing research.

8.3.1   Rely on Authoritative Sources

Editors are expected to cite content on Wikipedia from reliable sources such as books, journal articles, newspaper articles, Web pages, recordings, and other references. Nonetheless, curating content from different references can be a demanding task as information from different references might confuse editors and increase their level of uncertainty. Thus, editors usually rely on sources that originate from authorities such as the CDC and the WHO in order to clear any form of doubt or uncertainty. As reported in Table 8.4, relying on authoritative source is the most commonly used strategy in the talk pages of articles related to new disease outbreaks, as it accounts for around 37.5% of the total number of posts that contain the strategies to manage uncertainty.

Zika Virus:

As discussed earlier, Wikipedia editors experienced scientific uncertainty when they came across information about the relationship between the Zika virus and microcephaly. Thereupon, one of the editors cited the CDC to clarify this uncertainty in the Wikipedia article:

> *The lede was too strongly implying that there is a known causal link from Zika virus to microcephaly, so i made this edit (https://en.wikipedia.org/w/....) to clarify, using CDC as a source. The link is truly not known to be causal yet so implying that is not right for the article. The New York Times as of 12 February update (http://www.nytimes.com/interactive/....) concurs, as well, though they're not really MEDRS so i sourced to CDC. We need to watch that the article doesn't misrepresent what is known–15 February 2016 (UTC)*

Ebola Virus Disease:

Another editor cited the WHO while trying to resolve the uncertainty that emerged as a result of locating inconsistent scientific terminology to describe the virus that caused the Ebola virus outbreak:

> *Anyway, the WHO seems to use "Zaire species" when they give reports, should we stick with that?–2 November 2014 (UTC)*

Similarly, other editors avoided adding any information to the Wikipedia articles unless it was supported by the CDC or the WHO:

> *It is a primary source and uses a mathematical model. It has not been adopted by either the CDC or the WHO. Thus not convinced it is really notable–14 November 2014 (UTC)*

In some cases, statements from authoritative sources might also be perceived by the editors as uncertain. For example, an editor posted an inquiry regarding finding contradictory information about the transmission of the Ebola virus as cited by the CDC and the WHO. Another editor replied to his question stating:

> *Have you written to the CDC to ask them to clarify?–22 August 2014 (UTC)*

Swine Influenza:
Equivalently, a different editor reported finding inconsistent information about an antiviral drug used for treating swine influenza, so he received a reply stating:

> *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–27 April 2009 (UTC)*

Similar to the Ebola article, the editors tried to clear the doubts related to spotting inconsistent scientific terminologies by simply following the terminologies used by the WHO and the CDC:

> *Hey there. Yes, I do see your argument, but unfortunately with organisations like the CDC and WHO all referring to this as "swine inuenza" (see WHO for example (http://www.who.int/csr/....)) we have to do the same on Wikipedia. Once we have more information on this H1N1 strain we could split the articles into one on this human-transmissible strain and a second more general one about "swine inuenza", but presently we're stuck with this nomenclature–29 April 2009 (UTC)*

8.3.2   Report the Uncertainty in the Article

Editors could not succeed in finding reliable references to clarify different forms of uncertainty. In such cases, a strategy that they use is to report the uncertain statements in articles so as to explicitly inform the readers about these uncertainties. Reporting the uncertainty constitutes around 45% of the total number of posts that contain strategies

to manage the uncertainty in the articles related to new disease outbreaks and the articles related to chronic diseases combined. Reporting the uncertainty is also the most prevalent strategy used in the talk pages of the articles related to chronic diseases, as it constitutes around 58% of the total number of posts that contain strategies to manage uncertainty.

Zika Virus:

One of the editors argued that it was necessary to report that scientists had not yet confirmed that there was a link between Zika and microcephaly:

> *According to Scientific American (http://www.scientificamerican.com/article/....)*
> *and the New York Times (http://www.nytimes.com/interactive/2016/....)*
> *within the last two days, a causal link has not been definitively established.*
> *The article should reflect this and not make it seem conclusive when it's*
> *not. Therefore, i made this edit (https://en.wikipedia.org/w/index.php?....)*
> *to include this in the section–30 January 2016 (UTC)*

Furthermore, reporting an uncertainty could include finding phrases to tune down the language and to make the information sounds less definitive:

> *While the virus "may" spread from the mother to the baby during pregnancy*
> *I am not seeing evidence that this always occurs or that if it occurs it always*
> *results in severe problems. So toned it down a bit–15 April 2016 (UTC)*

Ebola Virus Disease:

During the Ebola outbreak, the editors also turned to using less conclusive statements to report the uncertainty in articles:

> *All my reading seems to indicate only mammals if anyone can find contrary*
> *please mention it. But 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–20 October 2014 (UTC)*

H7N9:

Suggesting less definitive statements to report the uncertainty is also a strategy that has been used by the editors of the H7N9 article:

> *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 ex-*
> *planations for this pattern have been proposed, as yet, the reason for this is*
> *not known–4 May 2013 (UTC)*

Diabetes:

Reporting different forms of uncertainty in articles related to chronic diseases has been adopted by editors as a strategy to manage uncertainty. For example, an editor suggested adopting it as a strategy to edit content concerning the relationship between testosterone deficiency and Type 2 diabetes:

> *The last two quotes comes from the conclusion. Emphasis mine: it seems apparent that there is considerable discussion and hypothesis on the thing, but no clear definitive consensus. It seems that what we can say is that: there is for sure some kind link between testosterone deficiency and T2D; androgen therapy has produced promising results in hypogonadal subjects but it is still uncertain if it is a solid therapy; the mechanism of the relationship is still being debated and while a lot of studies go in the same direction, the thing is far from being set in stone. I therefore, on this basis, support inclusion of the material in the article, in the following way:*
>
> - *Declare as considered factual the existence of some kind of relationship between testosterone and T2M, IR etc. as per studies on hypogonadal men.*
>
> - *Declare as still ongoing investigation the mechanism linking the two.*
>
> - *Declare as ongoing serious consideration, but still not proven, the possibility of testosterone therapy as a therapy for diabetes in hypogonadal subjects.–25 October 2009 (UTC)*

A different post revealed that the editors tried to be explicit about the current state of research on finding a cure for diabetes:

> *We often use the phrase the "cure is difficult" when it verges on impossible but very rare events do occur that we wish not to discount. Are you able to get a copy of the paper above on terminology of "cure". It just came out. We need to specifically state that the exotic diets and magics pills are unsupported by evidence however we have case studies of pancreatic transplants and studies of obesity surgery. Have you seen any guidelines / reviews that state that no cure exists?–9 November 2009 (UTC)*

Obesity:

Occasionally, editors create a section in Wikipedia articles that is dedicated to reporting cases of scientific uncertainty. For instance, one editor recommended adding uncertain content to the research section and reporting ongoing research on the treatment for obesity:

> *If this is added it would go in a section on Research rather than in the treatment section–15 July 2010 (UTC)*

Cancer:

Adding uncertain claims in the research section has also been suggested as a strategy to manage the uncertainty by the editors of the cancer Wikipedia article:

> *Yes, in the Research section, not in the lead. I think the sentence should also have "…but there is no conclusive evidence that any particular phytochemical prevents or fights cancer in humans"–15 July 2013 (UTC)*

Moreover, the editors of the cancer Wikipedia article discussed the possibility of reporting uncertain claims about the relationship between cancer and a healthy diet as well as between cancer and stress:

> *I actually agree with you.... in the suggestion above that the article say something like "the evidence for a specific factor is not definitive". The ACS seems to be saying, "Here is a healthy diet. It may even lower your risk of cancer." but there isn't good evidence demonstrating it–31 July 2013 (UTC)*

> *Under "Causes", second sentence, "stress" should not be listed as a cause of cancer. Most major cancer organisations (Cancer Research UK, American Cancer Society, Cancer Council Australia, etc) have made public statements saying that stress is not a known cause of cancer, with the American Cancer Society stating categorically that "No scientific evidence has shown that a person's personality or outlook can affect their cancer risk". It's a minor edit, but kind of important–29 November 2013 (UTC)*

Asthma:

Similar to other articles related to chronic conditions, the editors of the asthma Wikipedia article considered adjusting the language and the reporting of uncertain information about the possible treatments of asthma:

> *We could mention it (I have noticed that it's FDA approved), but perhaps in a "research directions" section. What we cannot suggest is that everyone with severe uncontrolled asthma is offered bronchial thermoplasty, because that is simply not the case–2 August 2011 (UTC)*

> *Have toned down the thermoplasty comments–20 January 2013 (UTC)*

8.3.3 Ignore the Uncertainty

In other cases, mentioning uncertain information may confuse the readers and cause the articles to become excessively complex. Thus, the editors might decide not to mention the uncertain information at all and, in effect, edit it out until more certain knowledge unfolds. As shown in Table 8.4, ignoring the uncertainty represents around 27% of the total number of posts that contain strategies to manage the uncertainty in articles related to both new disease outbreaks and chronic diseases.

Ebola Virus Disease:
The following post illustrates that it is challenging to find reliable literature to support a claim about the possible Ebola treatments. Therefore, not mentioning this treatment in the article might emerge as the best strategy to deal with this uncertainty:

> *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–11 July 2006 (UTC)*

In addition, ignoring the uncertainty also includes providing general information instead of directing the readers' attention to uncertain and possibly confusing details. The following post demonstrates this strategy to manage the uncertainty after an editor found inconsistent dates of the suspected first case:

> *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–31 October 2014 (UTC)*

H7N9:
Ignoring the uncertainty could be a temporary decision that the editors make until they have more certain knowledge and more references to rely on:

> *I think it would make sense to wait. After the outbreak dies down it will be easier to find a single trustworthy and reliable source that provides the necessary data. As it stands currently we'd probably have to assemble the data ourselves (verging on WP:OR). I also think it would be difficult to keep up to date (we have trouble as it is just keeping the counts for total cases and deaths up date, accurate, and well-sourced)–7 May 2013 (UTC)*

Another editor suggested to completely remove the number of people affected by the illness and the number of deaths from the article as a result of not finding sources that regularly update these numbers:

> *I'll remove the numbers–26 March 2014 (UTC)*

Diabetes:

Similar to the editors of the Ebola article, the editors of the diabetes article also agreed to be as brief as possible and avoid reporting uncertain and unconfirmed details in the article. This can be seen in the discussions related to the association between diabetes and chromium supplementation:

> *Proving a negative can be difficult. Whether or not chromium has an important role is still debatable, but it seems that any role it does play seems minor, and an overview article should reflect that and not give to much attention to a minor detail–24 July 2006*

Brevity is also observed in the discussions related to the effectiveness of specific surgery in treating diabetes:

> *Studies so far were carried out on only small number of patients and long term outcome is not clear. I think we would rather omit this at present or very brief at best–26 September 2007 (UTC)*

Obesity:

Like the editors of the H7N9 article, the editors of the obesity article decided to temporarily ignore the uncertainty around the potential treatments for obesity:

> *This is all fascinating, but modafinil is not registered for this indication, not recommended (yet) in any professional guidelines, and only one of a legion of drugs that has been investigated - from amphetamines to zonisamide. I strongly suggest we don't devote too much time to experimental data until well-designed large phase III studies with realistic end points have been conducted. Otherwise this article is likely to get flooded–30 May 2007 (UTC)*

> *If a drug is in phase III trials that might be appropriate, but in this case it's probably better to wait until there is some good data rather than proof of principle only–13 July 2010 (UTC)*

Cancer:

Ignoring the uncertainty has been proposed as a strategy to manage the uncertainty around the uncertain medical information related to cancer, such as around the role that DNA repair and apoptosis play in cancer:

> *Please understand that what we're presently presenting as solid truths are actually scientific controversies. These should NOT be elaborated on in depth over here. Let's stick with what is known–23 Jun 2005 (UTC)*

8.3.4 Consult Experts for Advice

As discussed, editors turn to the Wikipedia talk pages to ask other editors questions when they come across some uncertain information. On specific occasions, these questions are directed to domain experts such as doctors or health professionals, who are also involved in editing health-related articles. Consulting experts for advice is a strategy that has been used by the editors of the articles related to new disease outbreaks and chronic diseases.

Ebola Virus Disease:
Consulting an expert for advice has been used as a strategy to address the uncertainty regarding the immunity of Ebola survivors:

> *Does anyone know an immunologist so we can get this right?–6 October 2014 (UTC)*

Further, calling for experts' input has been used as a strategy to address the uncertainty concerning the transmission mechanism of the Ebola virus:

> *Zoologists and microbiologists please help – this is an excellent question that I think would enrich the article–14 October 2014 (UTC)*

A different editor emphasised that finding domain experts to clarify information is not always a straightforward task as the experts could also be unsure about certain dimensions of the diseases:

> *I put in a request for assistance at wikipedia project medical. I suspect that the experience and information is still, sadly lacking–7 October 2014(UTC)*

Diabetes:
Posting a request for help from a domain expert has also been perceived as a strategy to manage uncertainty with respect to the information about chronic diseases. For example, one of the editors posted a request to clarify scientific assertions related to the glucose generation in the body:

> *Again, can we have some clarification from someone who actually knows the biochemical pathways?–14 Feb 2005 (UTC)*

Cancer:
Similarly, editors ask for specialists' help to clear doubts about specific cancer-related information. This includes clarifying the meaning of specific terminologies such as 'benign' tumour:

> *Someone with more of a cell biology background might be able to give you a more in-depth answer–Feb 11, 2005 (UTC)*

8.3.5   Set Up a Mailing List about Ongoing Research

Other editors might take additional steps to identify and track the ongoing research about specific diseases. These steps involve setting up a mailing list to gather information and address different forms of uncertainty. This strategy has been observed in the talk pages of the Zika virus Wikipedia article:

> *I am setting up a mailing list (https://groups.google.com/....) 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–4 March 2016 (UTC)*

Articles	Authoritative Sources	Report the Uncertainty	Ignore the Uncertainty	Consult Experts	Mailing List
Zika Virus	3	3	0	0	1
Ebola Virus Disease	4	2	2	3	0
MERS	0	0	0	0	0
H7N9	0	1	3	0	0
Swine Influenza	2	0	0	0	0
Diabetes	0	9	4	1	0
Obesity	1	2	4	0	0
Cancer	1	8	2	1	0
Asthma	0	3	2	0	0
Cardiovascular Disease	0	1	0	0	0
New Disease Outbreaks	9	6	5	3	1
Chronic Diseases	2	22	12	2	0
Strategies	11	28	17	5	1

TABLE 8.4: Frequency of Strategies to Manage Uncertainty in Wikipedia Articles

In general, posts that contain strategies to manage uncertainty comprise only 4% of the total number of posts in the talk pages of the articles related to new disease outbreaks and chronic diseases. Furthermore, Figure 8.11 shows that editors can mention and suggest multiple strategies in the same posts to manage uncertainty. For example, it is likely that editors discuss reporting the uncertainty as well as relying on authoritative sources in the same post.

FIGURE 8.11: Overlap Between Different Strategies to Manage Uncertainty

8.4   Summary

This chapter focused on revealing the Wikipedia editors' response to the uncertainty during crises. This was done by distinguishing the forms and strategies used by Wikipedia editors to manage uncertainty during crises. To answer RQ4, scientific uncertainty, references uncertainty, and conflicting references are the key forms of uncertainty that emerge during crises. Further, to answer RQ5, relying on authoritative sources, reporting the uncertainty in the article, ignoring the uncertainty, consulting experts for advice, and setting up a mailing list about ongoing research are considered the main strategies to resolve the uncertainty in Wikipedia during crises. These results led to the development of a framework that models the forms and the strategies to manage uncertainty during crises. This framework is illustrated in Figure 8.12.
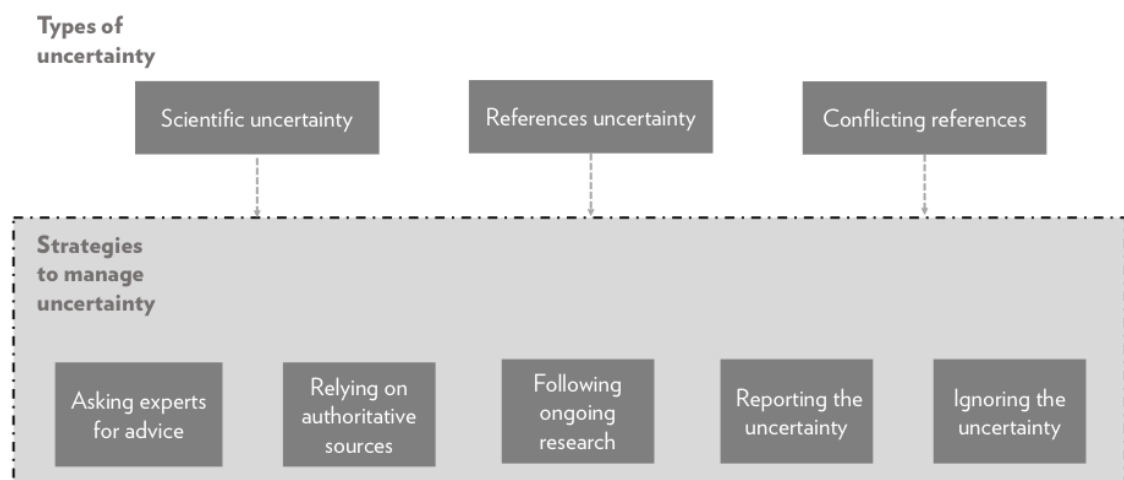
FIGURE 8.12: Uncertainty and Strategies to Manage Uncertainty on Wikipedia During Crises

Chapter 9

Overall Discussion and Conclusion

'Wikipedia is a precious artefact of the Web and a beacon of light in a world that is becoming increasingly blurred in terms of ethical values. We must all work hard to enable it to continue to flourish for the good of humanity.'

— Dame Wendy Hall
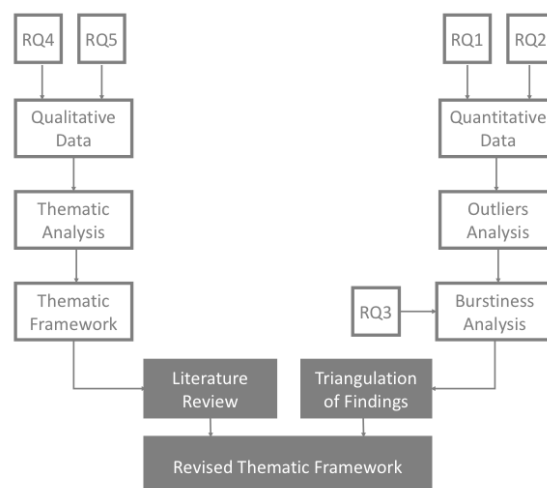
FIGURE 9.1: Chapter's Placement within Wider Research Design

This chapter presents the final discussions and conclusions based on the data collected and analysed in the preceding chapters. The previous chapter concluded by presenting a framework that models the forms and strategies to manage the uncertainty on Wikipedia during crises. The goal of this chapter is to discuss and triangulate the quantitative and qualitative findings and linking the findings to the literature and to previous theories. This discussion helped to gain a comprehensive understanding of the Wikipedia editors'

response to crises as well as to refine the thematic framework that models the uncertainty on Wikipedia during crises. The part of this work presented in this chapter was published as (Al Tamime et al., 2018b) and (Al Tamime et al., 2019). The limitations and future work opportunities are presented, before the final conclusions and overall contributions are made.

9.1 Triangulating Quantitative and Qualitative Findings

The section aims to discuss the findings that were reported in Chapters 6, 7, and 8, as well as to relate these findings to the literature and to theories of information uncertainty as presented in Appendix B. The following section also aims to refine the the designed framework illustrated in Figure 8.12 on the basis of the concept of social machines.

9.1.1 Quantitative Findings

This research's goal is to study Wikipedia editors' response to crises by discussing the quantitative findings. These discussions reveal patterns of editing behaviours during crises and identify and compare responses to the related articles. These discussions address the following three research questions: What are the characteristics of Wikipedia editorial activities during new disease outbreak crises?(RQ1); What are the differences and similarities between editorial activities on Wikipedia articles related to new disease outbreak crises and editorial activities on Wikipedia articles related to chronic diseases?(RQ2); and What are the main factors that influence Wikipedia editorial activities during new disease outbreak crises?(RQ3).

9.1.1.1 Model of Burstiness

The quantitative analysis on Wikipedia revision history suggests that there are similarities and differences between Wikipedia articles on new disease outbreaks and articles on chronic diseases. Outliers were detected in the dataset of the number of revisions and the number of editors in both new disease outbreak articles and chronic diseases articles. This led to the conclusion that articles related to new disease outbreaks as well as chronic diseases could receive sudden attention from editors. This sudden attention is not necessarily triggered by an external event such as a crisis. Nevertheless, a careful examination of burstiness by detecting outliers that occur simultaneously in the dataset of the number of revisions and the number of editors led to a different conclusion. In contrast to articles related to new disease outbreaks, articles related to chronic diseases rarely experienced burstiness. This emphasised that an external event such as a crisis could influence the editing activities on Wikipedia articles. In fact, external events could

also influence the magnitude of attention given to Wikipedia articles as the classification of burstiness in articles related to new disease outbreaks resulted in higher thresholds than the classification of burstiness in articles related to chronic diseases.

In the context of this research, the external event was the outbreak of a new disease. This event influenced the editing activities of Wikipedia articles because there was a lack of knowledge and uncertainty about the etiology of the disease, its spread, and its consequences for public health. The external event was a cause of concern not only to specialists in the respective fields but also to the general public. These findings motivated us to conduct an in-depth analysis of the editorial bursts on Wikipedia articles related to new disease outbreaks over time. Based on this analysis, a model of burstiness was developed. This model consists of the following stages:

Pre-existing Stasis $\longrightarrow$ Triggering Event $\longrightarrow$ Mediation $\longrightarrow$ Revisions $\longrightarrow$ Stasis

This model indicates that Wikipedia articles are generally static (i.e. they have a stable number of revisions over time). However, a triggering event may result in a sudden increase in editing activities or burstiness. Eventually, articles return to the stasis stage with a relatively stable or small number of editing activities.

In addition, articles of new disease outbreaks could experience high, moderate, or low burstiness. The intensity of burstiness was sensitive to the mediation of the triggering event. In particular, this study revealed that several factors mediate burstiness in Wikipedia articles; these are as follows: 1) the widespread and fast broadcast of the outbreak from news sources and trusted health agencies, such as the World Health Organization, which identified some diseases as public health emergencies that require a coordinated international response; (2) uncertainty and lack of knowledge about the cause of the outbreaks and the precise modes of disease transmission; (3) the mortality and morbidity threats of the outbreaks, for example, the Ebola virus had the threat of resulting in a mortality rate of 90%; and 4) the general public's perceived salience and anxiety about the potential spread of a deadly disease and about its other known or unknown consequences. Correspondingly, emotions (Gersick, 1991) play a role in amplifying the attention paid to a given Wikipedia article and in causing high burstiness in articles that are related to topics that the public mostly perceive as a high threat. These factors when combined help to explain why MERS (which has a mortality rate of approximately 50%) had lower burstiness than the Ebola virus (which has a mortality rate of 90%). Moreover, the intense news coverage of specific diseases, such as Zika, Ebola, and swine flu, tended to dramatize the threat and caused more anxiety among the general public (Klemm et al., 2016).

9.1.1.2 Theory of Punctuated Equilibrium

The theory of punctuated equilibrium can provide a well-substantiated explanation of this research's findings on burstiness in articles related to new disease outbreaks. The concept of burstiness has been recognised by the theory of punctuated equilibrium as "the history of evolution is not one of stately unfolding but a story of homeostatic equilibria, disturbed only "rarely" (i.e. rather often in the fullness of time) by the rapid and episodic events of speciation" (Gould and Eldredge, 1972). As such, the theory of punctuated equilibrium is closely aligned to explaining that a triggering event might cause a burst on a Wikipedia article. The theory is also aligned to explain how articles subside to relative stasis after burstiness. This is affirmed as "evolutionary change is characterised by species and individuals as homeostatic systems—as amazingly well-buffered to resist change and maintain stability in the face of disturbing influences" (Gould and Eldredge, 1972).

This research is closely connected to other research that explores the editing activities on Wikipedia following crises such as the Tokoku earthquake and a tsunami (Keegan et al., 2011). In addition, this research was connected to other research that applied the theory of punctuated equilibrium to understand an activity or an action on the Web, such as to understand how languished electronic petitions might suddenly, through an external trigger, encourage the initiation of political actions (Dumas et al., 2015; Margetts et al., 2016). However, this research extends previous work by analysing articles related to five different crises or new disease outbreaks. This research not only is limited to an analysis of burstiness as a proxy of the editing activities across articles, but also includes classifying burstiness as a way to compare the attention given to certain articles. Moreover, this research offers a longitudinal analysis of the editing activities before, during, and after new disease outbreaks. This has given the opportunity to apply the theory of punctuated equilibrium to expound the stage of triggering attention to an article as well as to expound the stage of returning to stasis.

9.1.1.3 Humans and Bots Role

Humans and bots play a role in editing articles during new disease outbreaks. Even though that bots play a trivial or no role in triggering burstiness during new disease outbreak crises, their edits are critical to the knowledge production process on Wikipedia. Bots are responsible to make many changes that would be tedious and take a long time for a human to do alone. In particular, bots can automatically protect the encyclopedia from vandalism, correct typographical errors, fix templates and citations, as well as add or change links on Wikipedia pages. This demonstrates that bots have a significant effect on the kinds of activities that are made possible on Wikipedia, especially in regard to fighting and reverting vandalism (Geiger, 2009). These humans and non-humans

work jointly to produce and maintain a social order that is required for the process of knowledge production on Wikipedia with the participation of diverse and geographically distributed contributors (Geiger, 2009).

9.1.2 Qualitative Findings

The quantitative findings discussed in the previous section helped to reveal and compare the editing activities on Wikipedia during crises or new disease outbreaks. These findings confirmed that editors' attention to the corresponding Wikipedia articles was influenced by an uncertainty and lack of knowledge about the diseases. This raised questions about whether the uncertainty during crises can also be experienced by Wikipedia editors while working on articles related to new disease outbreaks. However, given that uncertainty is a vague concept, the questions that persist are: What are the main types of uncertainty that emerge while editing Wikipedia articles related to new disease outbreak crises and Wikipedia articles related to chronic diseases?(RQ4); and What are the key Wikipedia editors' strategies for managing uncertainty on articles related to new disease outbreak crises and articles related to chronic diseases?(RQ5). To answer these questions, this section discusses and evaluates the qualitative results derived from Wikipedia talk pages.

9.1.2.1 Multiple Forms of Uncertainty

Multiple forms of uncertainty are expressed by Wikipedia editors during a disease outbreak. These forms emerge in Wikipedia talk pages as scientific uncertainty, conflicting references uncertainty, and references uncertainty. Scientific uncertainty is expressed by editors when scientific information is reported in references as uncertain or if relevant references are either absent or poor. However, references uncertainty is expressed by editors when trying to clarify claims within references, checking whether certain references are valid, or simply questioning the claims coming from some references. Moreover, conflicting references uncertainty is expressed by editors when coming across incompatible claims mentioned in certain publications as well as in the literature in general. These different forms of uncertainty can be categorised into scientific and editorial levels uncertainties. Scientific-level uncertainties are related to doubt and ambiguity concerning the scientific information about diseases. In contrast, editorial-level uncertainties are related to doubt and ambiguity concerning the references used to edit and build Wikipedia articles. Editors also had doubts about the information and references that belong to authoritative sources such as references that belong to international public health agencies (e.g. WHO and CDC). The different forms of uncertainty are often multi-layered and interconnected, leading to an elevation of uncertainty among editors. For instance, scientific uncertainty might also provoke references or conflicting references uncertainty as incomplete research and untested knowledge about diseases might make it difficult to find and clarify information from references during crises and disease outbreaks.

Similar to articles related to new disease outbreaks, editors also experienced scientific uncertainty, conflicting references uncertainty, and references uncertainty while editing articles related to chronic diseases. However, in relation to the total number of posts, the number of posts that include different forms of uncertainty tends to be slightly higher in articles related to new disease outbreaks. Moreover, an analysis of posts in articles related to new disease outbreaks revealed that editors doubt sources that belong to authoritative sources such as WHO and CDC and not only sources that belong to academic journals and papers, as is often the case in articles related to chronic diseases. This again pertains to the nature of the uncertainty that emanates during new disease outbreaks where there are limited publications and academic studies about diseases, which pushes people to rely more on updates from health agencies and press reports.

This research is interconnected with other research that explored how uncertainty is conveyed on social media and online forums in various contexts, including uncertainty that accompanies rumours (Starbird et al., 2016) and uncertainty that accompanies travel-related decisions (Gui et al., 2017). In particular, Starbird et al. (2016) described several milling behaviours—behaviours that involve gathering during a crisis to discuss and attempt to understand the cause of an event (Turner et al., 1957) —in rumoring Tweets during events related to crises. These milling behaviours include asking leading and open questions to either seek information or communicate doubt or uncertainty about the related information, challenging a rumour, discussing different forms of uncertainty and contradictory information about rumours, questioning the source of the rumours, and voicing different emotions such as incredulity, hope, and fear. Gui et al. (2017) found that people experienced uncertainty during the Zika outbreak as the information tended to be incomplete, inaccurate, and insufficient to make travel-related decisions. Mainly, posts on Reddit, TripAdvisor, and BabyCenter showed that uncertainty could emerge as a result of questioning and expressing distrust in authoritative sources, questioning whether the media are reporting information objectively, or there is overgeneralized and inaccurate reporting of information, as well as whether they are reporting inconsistent information from healthcare providers. (Qu et al., 2009) demonstrated that doubts on authentic reports were expressed and rumours were circulated in the Tianya online forum after the 2008 Sichuan Earthquake.

While existing research looked at the different forms of uncertainty expressed by the public during crises in different contexts, this research contributes to the existing literature by examining the forms of uncertainty expressed by Wikipedia editors while editing articles during public health crises. This research's findings agree with the existing findings that Wikipedia editors also find information to be incomplete, inaccurate, and contradictory during crises. In addition, Wikipedia editors question the information

and statements originating from authoritative sources. This is further explained and justified in the following section. The following section aims to identify the strategies that editors adopt to manage uncertainties during public health crises.

9.1.2.2 Strategies to Manage Uncertainty

Wikipedia editors use several strategies to manage uncertainty during a disease outbreak. These strategies include consulting and relying on authoritative sources, consulting experts for advice, reporting the uncertainty in Wikipedia articles, ignoring the uncertainty, and setting up a mailing list to send updates about relevant studies as they emerge over time. Each strategy can be used either alone or in combination with the other strategies.

Strategies that have been used to cope with uncertainty in articles related to new disease outbreaks tend to be similar to the strategies that have been used to cope with uncertainty in articles related to chronic diseases, with the exception of one strategy. Setting up a mailing list about ongoing research is a strategy that has been adopted in articles related to new disease outbreaks only. This is also related to the nature of new disease outbreaks, which are often characterised by a lack of scientific studies on the diseases, which in turn encourages editors to track the current research by setting up mailing lists.

Brashers (2001) delineated several strategies to manage uncertainty during an illness, including seeking information and adapting to chronic uncertainty. Information seeking is used to reduce uncertainty by striving to fill in the information gaps or knowledge gaps, or to confirm/refute the existing beliefs. Accordingly, this research's findings revealed that editors try to decrease the uncertainty during crises and disease outbreaks by seeking information from authoritative sources, by asking experts for advice, and by following the ongoing research on the disease outbreaks. Accepting uncertainty is another mechanism used by editors to adapt to an ongoing uncertainty. Adapting to a chronic uncertainty can involve ignoring the uncertainty-producing event in the interest of maintaining simplicity or amending planning and decision-making strategies (Brashers, 2001). Therefore, editors can sometimes adapt to an uncertainty by overlooking the uncertainty in the interest of maintaining the articles' simplicity.

Prior work focused on studying information-seeking practices to manage the uncertainty during crises and disease outbreaks. For example, Qu et al. (2009) discussed that that Tianya online forum members worked together to clarify rumours and clear doubts by seeking further information, cross-validating with other peoples' personal experience, as well as comparing and contrasting information with official announcements and news. Similarly, Vieweg et al. (2010) disclosed that various audiences at the individual and the community levels may seek information such as situational updates from Twitter during

crises and uncertain events. Gui et al. (2017) uncovered that people turned to online discussion forums to seek information that they could not otherwise obtain from authoritative sources, such as local information and alternative authoritative information. Different research also looked at strategies that were applied by the CDC to manage the uncertainty on Twitter during the Ebola virus outbreak (Dalrymple et al., 2016). Despite the fact that engaging and communicating with the public were recognised as strategies to manage uncertainty, the public did not look to engage back in the form of dialogue or conversation. Instead, the public were looking for concrete information to assess the potential risk factors. A subsequent study also investigated the use of Facebook for Zika-related outreach by the Ministry of Health (MOH) and the National Environmental Agency (NEA) in Singapore (Vijaykumar et al., 2017). The study found that posts related to disease investigation—such as posts that are related to the diagnosis of symptoms and other epidemiological surveillance activities—received the highest number of likes and shares on Facebook. This is likely to be driven by the public desire for constant updates about the Zika virus, coupled with the need to stay abreast of the situation (Vijaykumar et al., 2017).

This research's findings are aligned with previous findings, as Wikipedia editors tend to also seek information to handle an uncertainty by relying on experts for advice and on information coming from authoritative sources, and keeping a track of ongoing research. However, our study extends previous research that focused on seeking information only as a strategy to manage an uncertainty by exposing alternative strategies to adapt to the uncertainty during disease outbreaks. These strategies include ignoring the uncertainty and reporting the uncertainty in articles.

9.1.3   Quantitative and Qualitative Findings

In line with the methodology designed in Chapter 4, various pieces of this research were merged through triangulation to produce the final research outputs. These merged findings address the main research question: What effect do crises have on Wikipedia editing? A discussion of the triangulated findings is presented below as the key areas or themes that were elicited from the research process:

9.1.3.1   Integration of Quantitative and Qualitative Findings

This research aims to reveal unique characteristics of several forms of contribution to Wikipedia during crises. These forms are editing activities and posts to talk pages. Respectively, quantitative analysis is applied in order to examine change in editing activities on Wikipedia articles during new disease outbreaks crises. The results confirm that Wikipedia articles experience a sudden increase in the daily count of the number of revisions as well as the number of editors contributing to crises-related articles. This

increase is chiefly triggered by the surge of media coverage of the public health epidemics along with the WHO declarations that some epidemics are public health emergencies of international concern. This increase in editors' attentions to crises-related articles is also associated with lack of knowledge and uncertainty about the health epidemics. However, the concept of uncertainty is multifaceted and has not previously explored in the context of editing Wikipedia articles. This leads to observing posts to talk pages in order define different forms of uncertainty and strategies to manage uncertainty on Wikipedia during crises. The qualitative analysis revealed that Wikipedia editors have experienced scientific uncertainty, references uncertainty, and conflicting references uncertainty. This uncertainty has been managed by relying on authoritative sources, reporting the uncertainty in the articles, ignoring the uncertainty, consulting experts for advice, and setting up a mailing list about ongoing research. There are imperceptible qualitative differences between posts on the talk pages of articles related to infectious disease outbreaks and posts on the talk pages of articles related to chronic diseases. The percentage of posts that include forms of uncertainty in articles related to new disease outbreaks is higher than in articles related to chronic diseases. In additions, posts that include scepticism of references that belong to public health agencies such as WHO and CDC are essentially prevail in articles related to new disease outbreaks. Finally, posts that include setting up a mailing list about ongoing research as a strategy to manage uncertainty are exclusively prevail in articles related to new disease outbreaks.

9.1.3.2   Wikipedia as a Social Machine

As introduced earlier in this thesis, Wikipedia is a social machine that is 'co-created' by human and technological elements. Therefore, studying Wikipedia as a social machine is essential to capture the human dynamics of the system under consideration and broaden the way of thinking about participation (Shadbolt et al., 2019). This research demonstrates the value of 'co-creation' in social machines as Wikipedia editors participate to edit, correct, post, discuss, and manage uncertainty, and not just read or use Wikipedia. Additionally, social machines can work on a large scale as a large number of Wikipedia editors, who are located in different geographical locations, cooperate to solve problems in real time. In this context, a large number of editors gather on Wikipedia and use edits and talk pages to regularly update articles as well as to discuss scientific uncertainty, conflicting references uncertainty, and references uncertainty. As denoted, bots also play a role in editing articles during an uncertainty. This highlights that a technological component can participate in performing specific functions in social machines such as reverting vandalism, notifying users of mismatched brackets in recently edited articles, or archiving talk pages. In this regard, this research suggests several characteristics related to the significance of 'co-creation' during crises and disease outbreaks on Wikipedia, each of which is discussed below.

9.1.3.3   Wikipedia as an Evolving and Changing Network

The networks which underlie social machines are usually fluid, transient, and constantly evolving (Shadbolt et al., 2019). Network elements, whether they are human or artificial, are in constant flux as they are required to respond to changes that might be driven by internal factors as well as by exogenous factors that are beyond the social machines (Shadbolt et al., 2019; De Roure et al., 2015). The interest of network elements changes over time and may result in new elements to join or drop from the network in order to support the new agenda (Shadbolt et al., 2019). These characteristics of social machines have been highlighted in this research by looking at and detecting burstiness in Wikipedia articles related to new disease outbreaks. To illustrate the effect of crises on Wikipedia articles, the development of a model of burstiness was discussed in Section 9.1.1.1. This model confirmed that Wikipedia articles do not always remain in the stasis phase, and a triggering event might influence the attention given to particular articles and cause editors to intensely edit these articles. The number of edits by bots might increase as well, particularly in articles related to new disease outbreaks such as the swine influenza articles. Technological components that are not only limited to bots, but also include functions that are built in Wikipedia platforms (e.g. edit history and talk pages) were also a part of the Wikipedia network's response to the triggering events.

9.1.3.4   Learning from a Network on Wikipedia

Social machines can be divided into three different classes depending on the type and level of communication and engagement. Wikipedia belongs to the third class of social machines as not only are its activities social constituted and understood by the participants, but they also require engagement between participants (Shadbolt et al., 2019). This is because Wikipedia not only enables participants to edit content incrementally but also enables them to interact through conversations, debates, arguments, and even edit wars (Shadbolt et al., 2019). The role of the machine on Wikipedia is to facilitate the interaction between a group of people or communities of interest. The infrastructure that supports Wikipedia is sufficient to host the interactions as well as to reveal the current state of interactions in the revision histories and talk pages (Shadbolt et al., 2019). Consequently, this research was not only narrowed to studying edit histories by detecting outliers and burstiness, but also foregrounds several forms of interactions that take place in the Wikipedia talk pages. To delineate the effect of crises on Wikipedia articles, these forms of interactions have been highlighted in Section 9.1.2.2 where editors use the talk pages to express and discuss different forms of uncertainty during disease outbreaks. These uncertainties were also managed by using several strategies. As a result, interactions that were manifested in Wikipedia talk pages were essential to learn from a network during crises. Learning from a network includes discussing and aggregating feedback and opinions from different editors about different uncertain issues,

consulting with experts, and informing others about a mailing list to post any comments or questions.

9.2 Thematic Framework: Uncertainty in Social Machines During Crises

The thematic framework illustrated in Figure 8.12 presents various forms of uncertainty and strategies to manage the uncertainty during crises on Wikipedia. This section presents a refined framework based on the literature review on information uncertainty theories as well as based on the triangulation of the qualitative and quantitative findings discussed previously in this section. The refined framework consolidates Brashers's work on managing uncertainty during an illness by grouping different strategies to manage the uncertainty on Wikipedia under seeking information and adapting to the uncertainty. As discussed in Section 9.1.2.2, Brashers' theory of communication and uncertainty management has been mapped to the qualitative findings of uncertainty management on Wikipedia during disease outbreaks. In addition, the refined framework consolidates Shadbolt et al.'s work on social machines by grouping different strategies to manage the uncertainty on Wikipedia under learning from a network. As explained in Section 9.1.3.4, different characteristics related to social machines have been used to link and integrate the qualitative and quantitative findings related to editors' interactions to manage the uncertainty during disease outbreaks. As a whole, the refined framework of types and strategies to manage the uncertainty in social machines during crises is shown in Figure 9.2.
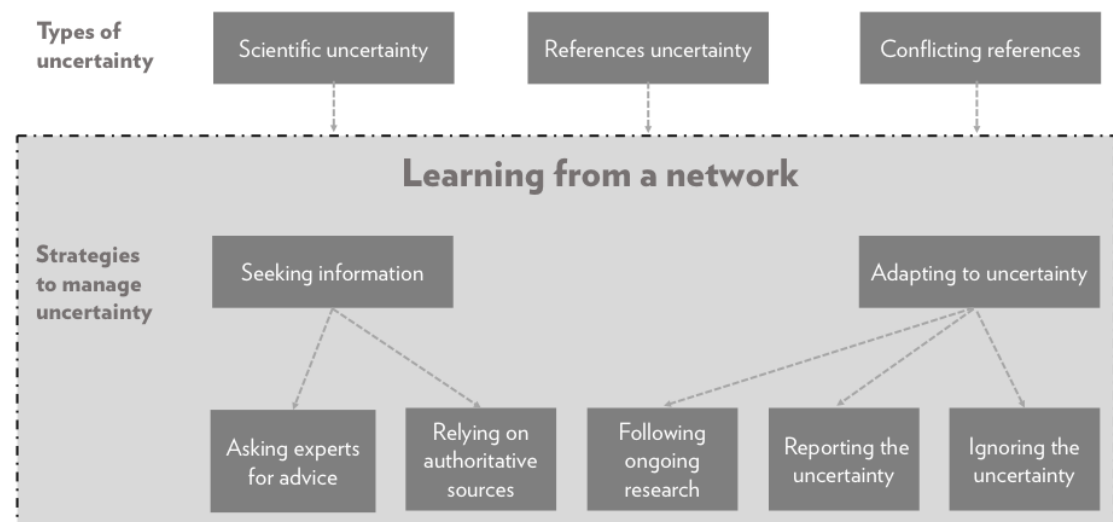
FIGURE 9.2: Uncertainty and Strategies to Manage Uncertainty on Wikipedia during Crises

9.3   Limitations and Future Work

The methodological approach used in this research mainly aimed to understand Wikipedia editors' reactions during uncertain events as well as to understand the types of uncertainties and strategies used by editors to manage these uncertainties. Primarily, even though that literature on trust and quality on Wikipedia has been reviewed in Appendix A, there has been no attempt to define or measure the trust or quality of Wikipedia articles during uncertain events. In addition, there has been no attempt to look at the relationship between different types of uncertainty and the quality of Wikipedia articles. Moreover, this work does not assess the references used to edit Wikipedia during crises or to look at the prevalence of citing a fake news website in articles. This includes looking at the prevalence of "rumours, conspiracy theories and lies deliberately shared in order to confuse, divert, frighten or manipulate the recipients" (Vijaykumar et al., 2019). However, this research provides suggestions for future research to expand the knowledge about the relationship between uncertainty, quality, and trust in Wikipedia articles. Trust and quality are considered a key challenge in building social healthcare machines (Shadbolt et al., 2019).

There are other clear limitations that characterise the design of this research and may have affected the interpretations of the findings. In particular, this work analysed articles written only in English, which could explain why some articles experienced low burstiness such as MERS. Furthermore, this work studied only specific cases of health-related articles. Further work may be needed to test the claims made about burstiness in articles related to new disease outbreaks in comparison to other types of articles such as articles related to chronic diseases. Moreover, the qualitative findings do not represent the views of all the Wikipedia editors; as not all editors post in talk pages. As shown in Table 8.2, only a subset of Wikipedia editors participates in talk pages, and hence, only this subset was scrutinized in this research.

Our results disclose that editors usually ask for experts' advice and rely on authoritative sources in order manage uncertainty during crises. This highlights the importance of further exploring the role of expertise, authority, and evidence-based medicine in constructing knowledge on Wikipedia. Even though that this thesis aimed to define uncertainty and strategies to manage uncertainty, this thesis did not discuss political influence or risk and attitude towards uncertainty. Moreover, even though that this thesis reviewed the literature on conflict and controversy in Chapter 2, this thesis did not study group dynamics including tribalism, negotiation, consensus, role of judgement, and conflict of interest. In particular, this thesis could include discussions about the relationship between editors with reference to anonymity, transparency, and responsibility. This thesis could also include discussions about governance and control of knowledge in

Wikipedia. This could be achieved by exploring the role of substantive experts in contributing health-related knowledge on Wikipedia. Substantive experts display extensive knowledge about a topic and provide substantive content to Wikipedia articles (Welser et al., 2011). These editors may invest more time and effort than other editors in order to fact check and discuss articles' content (Welser et al., 2011). These editors often try to gain attention by citing their real-world credentials and expertise on their Wikipedia profiles (Welser et al., 2011).

Finally, the qualitative research's findings demonstrated that there were posts that included both cases of uncertainty and cases of controversy—manifesting disagreement between editors—in Wikipedia talk pages (Appendix D). This work sets the future research agenda to further investigate the relationship between uncertainty and controversy. Potentially, this research aims to look at whether controversy can be detected by tracking phrases that express doubt and uncertainty on Wikipedia talk pages. Moreover, this research would benefit more from pursuing current research (Al Tamime et al., 2018a) that looks at bots' interactions with humans and the role that bots play during a controversy.

9.4 Overall Contribution and Conclusion

This thesis has asked a number of questions that aim to study the Wikipedia editors' response to crises. This thesis started by introducing Wikipedia as a socio-technical phenomenon. Then, this thesis proceeded to review the literature on crises and the crisis responses on Wikipedia. Paradoxically, previous research has focused solely on understanding the patterns related to viewing Wikipedia articles during crises. These studies have limited themselves to either the quantitative or qualitative paradigm. Accordingly, a gap in knowledge has been identified to affirm that there is a need to understand the process of Wikipedia articles' creation during crises. This is important to identify the impact of the element of swift attention on a socio-technical phenomenon such as Wikipedia. The research questions principally aimed to understand the effect of crises on Wikipedia articles, This thesis has answered the research questions by using quantitative and qualitative methods. Quantitative methods were applied to distinguish between editing patterns during crises. In addition, quantitative methods were used to study and recognise factors that could influence editing patterns during crises. Qualitative methods were applied to distinguish forms and strategies to manage the uncertainty during crises. The results showed that Wikipedia articles experienced burstiness during crises as identified by a sudden increase in both the number of edits and the number of editors per day. The level of burstiness was mediated by several factors including news coverage and WHO warnings that some diseases were major health threats. The results also showed that Wikipedia editors express uncertainty at the scientific and editorial levels.

Editors intend to manage uncertainty by consulting and relying on authoritative sources, consulting experts for advice, reporting the uncertainty in Wikipedia articles, ignoring the uncertainty, and setting up a mailing list. The findings were triangulated to gain a comprehensive view of the Wikipedia editors' responses to crises. The triangulation of the quantitative and qualitative findings was discussed by embracing the concept of social machines. This thesis's findings of editing patterns and talk pages' discussions to manage the uncertainty during crises demonstrated the importance of 'co-creation' in social machines. This was emphasised by unveiling several characteristics of social machines that acknowledge the effect of crises on Wikipedia articles. This was done by contemplating that Wikipedia is an evolving and changing network that allows editors to learn and engage with each other during crises. Consequently, a framework has been built to illustrate forms and strategies to manage the uncertainty on Wikipedia during crises.

This thesis offers a theoretical and methodological contribution to the field of Web Science:

1. **Conceptual Contribution:** This thesis builds a thematic framework that presents various forms of uncertainty and strategies to manage the uncertainty during crises on Wikipedia. This framework has been built using qualitative findings as guided by the thematic analysis. This framework was further refined by triangulating the quantitative and qualitative findings and by consolidating the information uncertainty theories. This framework was designed considering Wikipedia as a socio-technological phenomenon. This was done by studying both the social components covering the editors' responses to crises and the technical components covering the bots' responses to crises. This framework was further developed by considering Wikipedia to be a social machine. This was done by studying different characteristics related to the concept of 'co-creation' in social machines.

2. **Methodological Contribution:** The Web Science field is interdisciplinary by nature and requires a balance of methods and analytical techniques to answer Web Science research questions. This thesis benefited from using a mixed methods approach to address the research questions. Quantitative methods including outlier detection were utilised to measure the burstiness in editing activities. An assessment of different outlier detection methods was presented to select the most accurate outlier detection model. Qualitative methods including the thematic analysis were applied to divulge forms and strategies to manage the uncertainty in Wikipedia. These methods combined helped to understand closely the Wikipedia editors' responses to crises.

Appendix A

Trust on Wikipedia

A.1 Trust on Wikipedia

Can Wikipedia Be Trusted? Even though Wikipedia remains a useful and quick-to-access source, whether information on Wikipedia can be trusted is still open to question (Fallis, 2008). Ford et al. (2013) described the most commonly cited sources on Wikipedia and categorised them on the basis of the reliability of the sources, irrespective of whether they are primary, secondary, or tertiary sources[1]. The study found that primary sources, such as data and statistics, make a significant portion of citations on Wikipedia. Furthermore, primary sources are likely to be the most persistent and resistant to deletion by editors. Nevertheless, citations that are frequently cited on Wikipedia at a given point of time are not necessarily the ones that persist over time (Ford et al., 2013). Studies have discussed several concerns about the reliability of content on Wikipedia (fal; Halavais and Lackaff, 2008) and have argued that the reliability of Wikipedia is comparable to that of traditional encyclopaedias such as Britannica (Giles, 2005). Others have defended Wikipedia by suggesting that it should not be judged as an accurate body of published content existing at one moment in time, but rather as the embodiment of a continuous editing process that depends on the efforts of both amateurs and experts alike in order to improve entries (Niederer and van Dijck, 2010). Similarly, others have tried to explain that Wikipedia is more easily accessed, can be updated faster, covers more topics, and receives more contributions than traditional encyclopaedias (Magnus, 2009).

Building on different definitions and measures of trust, a large group of studies have attempted to explore different approaches to measure the trust in Wikipedia. The next section will discuss three suggested approaches: persistence and editors' reputations, metadata properties, and citation quantity.

---

[1]https://en.wikipedia.org/wiki/Wikipedia:Primary_Secondary_and_Tertiary_Sources

A.1.1 Persistence and Editors' Reputations

Trust based on content persistence proposes that trustworthy content is content that endures across revisions and editors' reviews of an article. Accordingly, different approaches and algorithms have been proposed to reveal and highlight persistent content on Wikipedia. Likewise, trustworthy editors are editors who write content that endures across the revisions and the reviews of an article. Thus, different approaches and algorithms have also been proposed to identify editors who have created persistent content on Wikipedia. These approaches help to build the reputation of editors on the basis of the persistence of their edits. Adler and de Alfaro (2007) introduced an algorithm that aims to compute the trust values of a given text in Wikipedia articles. This algorithm relies on a text tracking system that measures the text life on Wikipedia. Text life refers to how much of the text inserted by A is still present after B's edits. Furthermore, text life refers to how much of text reorganisation (i.e. text re-ordering and deletion) performed by A is still preserved after B's edits. Editors gain a reputation when their contributions prove to be long-lived, and their reputation suffers when their contributions are reverted or undone. Cross (2006) introduced an algorithm that labels text by colours in Wikipedia on the basis of its persistence. Along the same lines, Adler et al. (2008) implemented an algorithm that computes the trustworthiness of a word in an article on the basis of the reputation of the original author and the reputation of all the authors who edited text within a specific proximity of the word. The trust values of different words were revealed by highlighting the text using various background colours. Wöhner et al. (2011) introduced a novel method that depends on determining editors' reputations on the basis of the persistence of their edits. This method mainly aimed at establishing different profiles on the basis of the editors' behaviours and reputations so as to recognise malicious editors (e.g. editors who intentionally damage Wikipedia articles).

A.1.2 Metadata Properties

The research under this category relied on Wikipedia articles' content-exclusive properties and content-inclusive properties to measure the articles' trustworthiness. Content-exclusive properties include timestamp, article title, and summary. Content-inclusive properties include article length, number of images, and number of external links (West et al., 2012). Twidale (2005) used Wikipedia articles' metadata to construct a information quality metric. This information quality metric covers information authority and reputation (number of unique editors, total number of edits, connectivity, number of reverts, number of external links, number of registered, and anonymous user edits), completeness (article length, number of internal broken links, and number of internal links), complexity (readability scores), informativeness (information noise, diversity, and number of images), consistency (number of admin's edit shares and age), and currency

and volatility (median revert time). Dondio et al. (2006) developed two metrics that can be used to infer the articles' trustworthiness. These metrics were the users' distribution/leadership and stability. The users' distribution/leadership was calculated on the basis of the average number of edits per user, the standard deviation of edits, the percentage of edits produced by the most active users, the percentage of edits produced by users with more than n edits for a particular article, and the number of edits in the talk pages. The stability was calculated on the basis of the rate of edits from time t, the percentage of text different from the current version, and the version at time t. Blumenstock (2008) argued that word count is a simple metric that can be used to measure the quality of Wikipedia articles. Furthermore, relying on the word count is considerably more accurate than the complex methods proposed in the related work. Suh et al. (2008) designed WikiDashboard, which is a tool that uncovers Wikipedia editing patterns such as total edits, weekly editing trends, list of editors, and percentage of edits per user. Chevalier et al. (2010) worked on visualising Wikipedia metadata that could be used to reflect the trustworthiness of articles. These metadata include word count, number of contributors, number and length of edits, number of references, number of internal links, and length of discussions.

A.1.3 Citation Quantity

The PageRank algorithm is an algorithm that is used by search engines to rank pages by counting the number of links directed to a page. The PageRank approach has been applied by McGuinness et al.(2006) to rank the trustworthiness of Wikipedia articles. This is done mainly by calculating the number of links that exist internally between articles on Wikipedia. The 'internal PageRank' algorithm works by determining whether a concept or term in a Wikipedia article is linked (cited) to its associated article. This approach assumes that missing citations mean that non-cited articles are not trustworthy. Therefore, the numbers of citations and non-citations of an article are summed to determine the trustworthiness of an article. The trustworthiness of an author is the sum of the trustworthiness of all the articles that he/she has edited or created. The trustworthiness of an author is calculated by using a weighted average, taking into consideration the size of the articles. Similarly, other researchers suggested using a social network analysis of the articles-contributors network so as to find trustworthy articles. Social network analysis is "a research paradigm which tries to unravel patterns of social relationships across various individuals in a social context" (Korfiatis and Naeve, 2005). Wikipedia has two types of social networks: the articles' network and the contributors' network (Korfiatis and Naeve, 2005). The articles' network indicates that Wikipedia articles contain a link (citation) to other articles. The contributors' network indicates that Wikipedia articles have multiple contributors. Contributors might belong to the same articles or be interconnected, as common contributors work on two or more articles

in the same domain. After constructing an articles-contributors network of articles under WikiProject Medicine, Kane (2009) confirmed that higher-quality articles[2] (marked as Featured, A-Class, or Good articles by Wikipedians) have higher degree centrality and eigenvector centrality (degree centrality captures how well an article is connected to other articles via editors, while eigenvector centrality captures how well an article is connected to other well-connected articles). Well-connected articles are likely to contain more information and knowledge. Roy et al. (2013) proved that high-quality articles (marked as Featured articles by Wikipedians) have a lower clustering coefficient and average path length than low-quality articles. This indicates that high-quality articles are relatively centrally placed in the articles-contributors network, so they connect several articles via shared editors that would not otherwise be connected. Therefore, these articles are hubs that benefit from having access to a large number of contributions (Roy et al., 2013). In a later study, Kane and Ransbotham (2016) weighted the degree centrality by the quality of the Wikipedia articles under WikiProject Medicine (marked as Featured, A-Class, or Good articles by Wikipedians). Therefore, these weighted values ensure that contributors transfer better knowledge, not just more knowledge. Again, the study confirmed that weighted degree centrality and eigenvector centrality are positively related to the articles' quality.

---

[2]https://en.wikipedia.org/wiki/Wikipedia:WikiProject_Medicine/Assessment

Appendix B

Information Uncertainty Theories

This appendix summarises the existing literature on theories related to information uncertainty.

B.1   Definition of Information

Scholars have attempted to define, explain, and theorise the concept of *information*. The term 'information' can be defined in different ways such as information-as-process, information-as-knowledge, and information-as-thing (Buckland, 1991). Information-as-process is the act of informing or communicating of knowledge or news or the act of telling or being told of something. Information-as-knowledge is the process of increasing and decreasing uncertainty by using particular facts, subjects, or events. Information-as-thing is the act of using information such as data and documents as informative objects (Buckland, 1991). Nevertheless, there is no uniform definition of information. For example, Shannon (1948) conceptualised information as an eliminated uncertainty and focused on quantitatively analysing the transmission of information via a channel as well as measuring the amount of information in a message. Moreover, Ashby (1965) defined information as the change degree of things. In addition, Changli (1985) described information as an attribute of reflected things, whereas Chenguang (1989) referred to information as a reflected specialty (Zhiyong et al., 2007). Information cannot be evaluated without the awareness of the readership, authorial, and message contexts (Madden, 2000). The readership context is the context in which a message is received and interpreted by a mechanism, an organism, a community, or an organisation. The authorial context is the context in which the message originates, while the message context is the means by which information is transmitted including written, spoken, and facial expressions (Madden, 2000). The concepts of *information* and *uncertainty* are highly connected as uncertainty can be a "manifestation of some information deficiency, while information is viewed as the capacity to reduce uncertainty" (Klir, 2005).

Research on information uncertainty has been approached from different perspectives. Mainly, the literature on information uncertainty can be classified as mathematical and non-mathematical studies.

B.1.1   Mathematical Studies on Information Uncertainty

Different studies have aimed to quantify the information uncertainty by using different mathematical and statistical models. Accordingly, the information uncertainty can be studied as both aleatory and epistemic uncertainty. Aleatory uncertainty refers to the uncertainty that cannot be eliminated or reduced by collecting more knowledge or information (Li et al., 2013). For example, the outcomes of tossing a coin remains uncertain and unpredictable all the time irrespective of how many times the coin is tossed. Therefore, aleatory uncertainty derives from an inherent randomness in the behavior of the system under study (Helton et al., 2010). Aleatory uncertainty is also referred to as natural variability, objective uncertainty, external uncertainty, random uncertainty, stochastic uncertainty, inherent uncertainty, irreducible uncertainty, fundamental uncertainty, real-world uncertainty, or primary uncertainty (Li et al., 2013). The probability theory is considered one of the most widely used theories to study aleatory uncertainty. It is concerned with studying random events and determining the outcomes of these events. This is done by conducting repeated trials and event analyses to analyse the occurrence of events (Li et al., 2013; Ross, 2006). Accordingly, different methods and theories were developed on the basis of the probability theory. These include the Monte Carlo method, the Bayesian method, and the Dempster-Shafer theory. The Monte Carlo method solves mathematical problems by simulating random variables to approximate exact values (Sobol, 2018). In contrast, the Bayesian method provides a tool to make inferences by using an underlying graphical structure and the probability calculus (Holmes, 2008). Moreover, the Dempster-Shafer theory allows finding the confidence interval, containing the exact probability, by combining evidence from different sources and arriving at a degree of belief (Li et al., 2013; Klir, 1995). Shannon's information theory (Shannon, 1948) was also developed on the basis of the probability theory to measure and characterise the uncertainty in various modes and communication systems (Liang and Qian, 2008). According to Shannon, a general communication system consists of five parts (Lombardi et al., 2016):

- A source S which generates the message to be received at the destination.

- A transmitter T which turns the message generated at the source into a signal to be transmitted. The information can be encoded and implemented.

- A channel CH which is the medium used to transmit the signal from the transmitter to the receiver.

- A receiver R which reconstructs the message from the signal.

- A destination D which receives the message.

The core concepts of Shannon's theory are *uncertainty* and *entropy*. Entropy is a "measure of the uncertainty in a random variable and quantifies the expected value of the information (in bits) contained in a message" (Svtek, 2015). In other words, uncertainty and entropy are used to measure the information content of a message and the anti-information content of *noise* (Cole, 1993). These concepts realise that the communication signal is separate from its content (Albers, 2012). Thus, information and uncertainty can be presented as a paradox from the perspective of the sender (equation B.1) and from the perspective of the receiver of the message (equation B.2) (Cole, 1993):

$$\text{information} = \text{uncertainty} = \text{entropy} \tag{B.1}$$

$$\text{information} = \text{reduction of uncertainty/entropy} \tag{B.2}$$

Epistemic uncertainty derives from a lack of knowledge and the ability to measure and model the physical world (Helton et al., 2010). In contrast to aleatory uncertainty, uncertainty can be reduced or eliminated through increased understanding or research, or by obtaining more relevant data and knowledge (Swiler et al., 2009). Epistemic uncertainty is also called knowledge uncertainty, subjective uncertainty, internal uncertainty, incompleteness, functional uncertainty, informative uncertainty, or secondary uncertainty (Li et al., 2013).

Different theories have been used to represent epistemic uncertainty, including probability theory, fuzzy theory, info-gap theory, and possibility theory. The fuzzy theory addresses the issues of imprecision and vagueness inherent in various modes of approximate reasoning. This includes vague linguistic labels with ill-defined foundries such as young, good, and cold (Dai and Xu, 2012). Unlike the classical set theory which assesses whether an element either belongs to or does not belong to a set, the fuzzy theory permits the gradual memberships of elements to a set by assigning membership grades in the unit interval [0,1] (Li et al., 2013; Klir, 2004). An info-gap is a disparity between what the decision maker knows and what he could know (Ben-Haim, 2004). An info-gap model of uncertainty is a group of nested sets. Each set corresponds to a particular degree of knowledge deficiency, depending on the different levels of nesting. Each element in a set represents a possible event such as a possible model of a physical or social system (Ben-Haim, 2004). Similarly, the possibility theory aims to look at the possibility of evaluating whether an element belongs to a determinate set (Galindo, 2008). The theory discards or includes elements in a set by obtaining the relevant information and conducting experiments to gather evidence and decide which alternative element can be a member of a given set (Klir, 2004). Even though that uncertainty is classified into

aleatory uncertainty and epistemic uncertainty, it is difficult to set boundaries between the two types as they could be handled using the same methods and theories (Li et al., 2013). Other researchers have worked on developing a broader treatment of uncertainty-based information that is not restricted to a specific theory or framework. In particular, the Generalized Information Theory (GIT) aims to handle the information uncertainty by not committing to a particular uncertainty theory, but by selecting a theory that allows one to express ignorance and at the same time does not allow one to ignore any available information (Klir, 2005). Although the GIT acknowledges that finding a broad framework to deal with any type of information might be a long-term goal that might not be fully achieved, the theory tries to a design a blueprint that encompasses a broad spectrum of special mathematical areas to formalise the various types of uncertainty (Klir, 2005).

B.2   Non-mathematical Studies on Information Uncertainty

Various non-mathematical studies have been conducted to understand and manage information uncertainty. These studies were typically derived from the field of Communication Science. These theories include Uncertainty Reduction Theory, Predicted Outcome Value Theory, Problematic Integration Theory, Theory of Motivated Information Management, and Uncertainty Management Theory. The Uncertainty Reduction Theory, developed in 1975 by Charles Berger and Richard Calabrese, defines uncertainty as possible alternative predictions possible when interacting or meeting others for the first time. A greater number of perceived alternatives produce a greater sense of uncertainty and a stronger drive to reduce uncertainty (Bradac, 2001). This theory delineates three information-seeking strategies that people use to reduce uncertainty. These strategies are passive information-seeking, active information-seeking, and interactive information-seeking. Passive information-seeking involves observing how the other person responds or behaves in a given uncertain situation in order to obtain insights about him/her. Active information-seeking strategies entail taking action without directly interacting with the other person in a given uncertain situation such as by asking a third party about the person. Finally, interactive information-seeking strategies involve communicating directly with the person in order to reduce uncertainty such as by asking questions (Knobloch and McAninch, 2014). Alternatively, the Predicted Outcome Value Theory, introduced in 1986 by Michael Sunnafrank, assumes that the ultimate goal in initial interaction is the maximisation of the relational outcomes. The theory acknowledges that individuals seek information because they attempt to forecast the outcome of communicating with the other person whether it is a positive or negative outcome. These forecasts guide individuals to decide how and whether to continue with the interaction and the relationship with the other acquaintance (Sunnafrank, 1990).

Problematic Integration (PI) Theory, introduced by Austin Babrow in 1992, is concerned with situations in which forming and integrating (or synthesising) knowledge and values is problematic. These problematic situations include uncertainty, ambivalence, and clashing expectations and desires (Babrow, 2015). This theory proposes that people hold both probabilistic and evaluative orientations when trying to assess a given uncertain situation. These probabilistic orientations represent judgments that provide individuals with a systematic understanding of their environment. The probabilistic orientations assume that people follow certain beliefs and expectations with varying degrees of confidence, assurance, and certainty. In addition, the evaluative orientations refer to the value-laden assessments that individuals consider when evaluating how a given situation will affect their well-being. Evaluative orientations could span positive, negative, and neutral judgments. Probabilistic and evaluative orientations are interlinked and used dynamically to evaluate and make sense of a particular situation (Knobloch and McAninch, 2014).

The Theory of Motivated Information Management (TMIM) is a theoretical framework that focuses on how individuals manage uncertainty when facing important or challenging problems. This theory aims to understand why individuals seek or avoid information by looking at the gap between the level of uncertainty that a person possesses and the level of uncertainty that a person wishes to have. According to this theory, the process of information management folds into three phases: the interpretation phase, the evaluation phase, and the decision phase (Morse, 2014). The interpretation phase begins when individuals recognise a gap between the amount of uncertainty they have and the amount of uncertainty they desire. This gap can be described as the uncertainty discrepancy and is highly dependent on the emotional response and the level of anxiety surrounding an issue. In the evaluation phase, individuals assess the advantages and disadvantages of alternative information management strategies, such as by using a cost-benefit analysis. During this phase, individuals hold different expectations such as process-based expectancies and results-based expectancies. Process-based expectations are the outcomes that individuals expect from enacting a strategy. This means that individuals try to anticipate whether the process of seeking information may present costs and rewards irrespective of the information gained. In contrast, results-based expectancies are the consequences that individuals forecast from receiving information including the costs and rewards of the information gained. In the decision phase, individuals actually adopt an information management strategy that could include information seeking, information avoidance, and cognitive reappraisal. Information seeking includes actively looking for sources and opinions. Information avoidance includes actively or passively avoiding information. A cognitive reappraisal of the situation includes constantly reassessing the emotional response, the uncertainty discrepancy, and the adopted strategy (Knobloch and McAninch, 2014).

Dale Brashers (2001) developed the Uncertainty Management Theory (UMT) to understand how individuals experience and cope with uncertainty. Even though this theory acknowledges the complex and variable nature of uncertainty, uncertainty is defined as an inability to make sense of, assign value to, or predict the outcomes of events (Kosenko, 2014). This theory focuses on studying uncertainty in an illness as patients might experience uncertainty associated with the diagnoses, treatment, and ways of living with certain diseases (Brashers, 2001; Brashers and Hogan, 2013). According to Brashers, individuals gauge the uncertain events on the basis of the relevance and impact to their lives. Therefore, uncertainty appraisals are usually shaped by different types of emotional responses such as negative, positive, and neutral emotional responses. Negative emotional responses could be fear, anxiety, or even panic when an uncertainty is viewed as a danger or threat. However, positive emotional responses could be hope and optimism when an uncertainty is viewed as beneficial. Moreover, neutral emotional responses might involve feeling both negative and positive emotions or feeling indifference such as when the uncertainty is considered to be both a threat and an opportunity (Brashers, 2001).

Despite the fact that the process of managing uncertainty is multi-layered, interconnected, temporal, and likely to vary across contexts, Brashers highlighted four strategies to manage uncertainty. These are seeking and avoiding information, adapting to chronic uncertainty, obtaining assistance with uncertainty management through social support, and learning skills to manage uncertainty in general (Rains and Tukachinsky, 2015). Seeking information helps individuals reduce uncertainty by finding alternatives during uncertainty, adding knowledge that they lack, or confirming or refuting their current state of belief (Brashers, 2001; Rains, 2014). Seeking information, however, can also increase the uncertainty as information could emerge as unknown alternatives or as contradicting to the existing knowledge. Besides, avoiding information could be used as a strategy to maintain or increase uncertainty by shielding individuals from information that is usually overwhelming and distressing. Avoiding information could be carried out by individuals either intentionally or unintentionally. The second strategy is adapting to chronic uncertainty, which aims to help people live with the ongoing uncertainty in life such as to help people living with chronic conditions. Adapting to chronic uncertainty involves an adopting attitude and certain behaviours that aim to tolerate and appreciate the uncertainty. These behaviours include ignoring the uncertain event, redefining plans, and developing structure and routines around uncertain events. The third strategy is managing uncertainty through social support. Social support can help to manage uncertainty by gaining access to messages, opinions, and help in seeking, avoiding, or evaluating information. Meanwhile, social support can also increase and decrease uncertainty as the information gained from social circles can also increase or decrease uncertainty. The last strategy admits that managing uncertainty can be a complex process, as information can be challenging to manage and interpret or information

can be contradictory and inconsistent. Thus, individuals need to learn certain skills such as to be vigilant when the information is not easily obtained, to assess the credibility of information, and to constantly judge the information from social support groups. Overall, Brashers' Uncertainty Management Theory expands the previous theories on information uncertainty by discussing alternative strategies to manage uncertainty that are not limited to seeking or avoiding information.

Uncertainty Management Theory (UMT) has been applied to explore different topics related to uncertainty, mostly topics related to uncertainty during an illness. To be more precise, the UMT has been built on the basis of a focus group study that aims to explore how patients with HIV and AIDS manage and live with uncertainty associated with chronic illness (Brashers et al., 2000). Correspondingly, Brashers has looked at two cases to investigate the uncertainty during an illness. These cases include people living with HIV and genetic screening for the disease risk. The study revealed 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). These strategies have been incorporated to manage the uncertainty in information-retrieval systems. In addition, the UMT has been applied to explore the uncertainty of mothers at risk for, or carriers of the cancer gene by conducting a thematic analysis of 16 genetic counseling sessions between practitioners and mothers (Fisher et al., 2017). Moreover, the UMT has been applied to examine the uncertainty of people who experienced parental divorce when they were older than 18 years (Mikucki-Enyart et al., 2018). In the context of studying the Web, the UMT has been used as a theoretical framework to analyse the content of messages posted to an online community by the caregivers and parents of children with conditions related to clubfoot. The study found that the caregivers managed the uncertainty through information seeking such as by asking questions, self-disclosure, offering candidate answers, second-guessing, and passive information seeking (Oprescu et al., 2013). Furthermore, the UMT has been used to explore information-seeking behaviors when using the World Wide Web to attempt to manage forms of uncertainty that are related to skin cancer. Information seeking was conceptualised as a means for decreasing or increasing one's (actual) uncertainty and thereby resolving the discrepancy between one's actual and desired levels of uncertainty. The study affirmed that exposure to some types of information, such as information about severity and susceptibility of the disease, was systematically related to the information seekers' relative success at managing their uncertainty (Rains and Tukachinsky, 2015). The UMT was applied to offer a theoretical explanation for how individuals used the World Wide Web to acquire the health information related to cancer prevention. The study found that the relationship between the desired and the actual uncertainty was expected to be stronger among individuals who used the Web to seek information. These findings also provided evidence that the use of the Web for information seeking facilitates uncertainty management and appraisal (Rains, 2014).

Figure B.1 summarises the discussed information uncertainty theories.
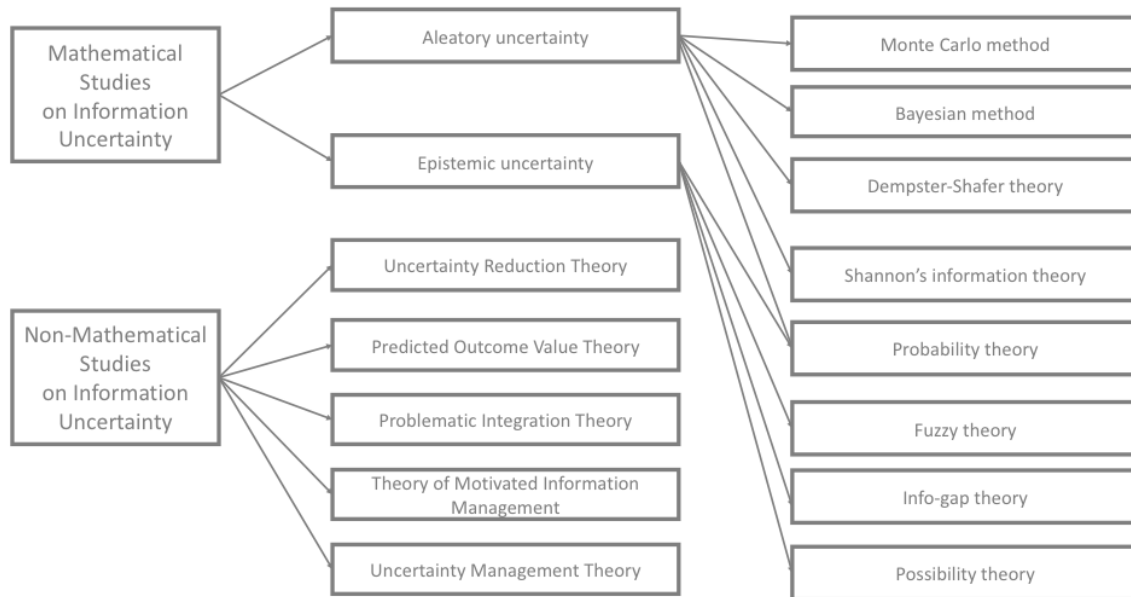
FIGURE B.1: Information Uncertainty Theories

B.3   Punctuated Equilibrium

In order to understand the uncertainty during periods of change and instability, this section presents the Punctuated Equilibrium theory. In 1972, paleontologists Stephen J. Gould and Niles Eldredge argued that the biological evolutionary change was not always 'slow and steady', as outlined by Charles Darwin in 1859 (Darwin, 1859). Instead, the periods of gradual evolution were *punctuated* by sudden revolutionary periods of rapid change (Eldredge and Gould, 1972; Van De Ven and Poole, 1995). Accordingly, the Theory of Punctuated Equilibrium (PE) claims that most of the evolutionary change occurs during relatively short bursts of activity, punctuated by long eras of evolutionary stasis (Woodberry et al., 2009). Besides biological evolution, many natural phenomena evolve intermittently with periods of tranquillity interrupted by bursts rather than by following a smooth gradual path. These phenomena include earthquakes, volcanic eruptions, solar flares, and gamma-ray bursts (Bak and Boettcher, 1997). Figure B.2 shows a comparison between the gradual and the punctuated biological evolutionary changes. The punctuated equilibria illustrate that evolutionary change is gradual, progressive, and possibly inevitable.

The punctuated equilibrium paradigm consists of three main components: deep structure, equilibrium periods, and revolutionary periods (Gersick, 1991). Deep structure is considered the basic structure into which units are organised and coexist with the basic
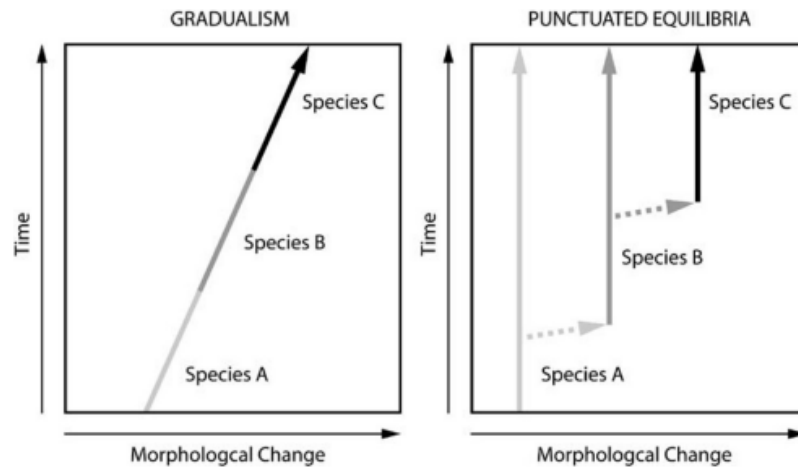
FIGURE B.2: 'Darwinian' View of the Gradual Evolution of Species (Left) Compared
with the 'Punctuated Equilibria' Concept of Stasis and Speciation
(Eldredge, 2008)

activity patterns that maintain a system's existence. Deep structures are determined
by a set of fundamental choices that make its parts highly stable. Defining these sets
of choices is a prerequisite for building a theory around punctuational models and can
vary from one system to another. These choices are also carried out and maintained
during equilibrium periods to ensure that the system's basic activity patterns stay the
same. However, systems during equilibrium periods can also make incremental adjust-
ments to compensate for internal or external perturbations without changing their deep
structures. The deep structure generates a strong inertia that prevents the system from
finding alternatives outside its own boundaries and pulls any deviations that do occur
back into line (Gersick, 1991). However, if the deep structure gets disrupted, the sys-
tem might go through a revolutionary period. Within this period, the system becomes
temporarily disorganised and undergoes a revolutionary change. Systems might become
disrupted as a result of (1) internal changes that include actions out of alignment with
each other or the environment, and (2) external or environmental changes that threaten
a system's ability to obtain resources. Internal changes usually occur when units or
members within systems initiate their own transitions and determine the end of a spe-
cific equilibrium period. In contrast, external or environmental changes can entail crises
and the subsequent attraction of newcomers to intervene and end an equilibrium period.
As a result, a revolutionary change involves a period of uncertainty where the old equi-
librium decays before getting crystallised into a new deep structure (Gersick, 1991).

Systems in transition are unpredictable, as they have the option to follow different
equally possible paths. In contrast to the relative predictability of equilibrium condi-
tions, neither the mechanics of human cognition nor the system itself absolutely 'dictates'
the outcome of a transition. In addition, the unpredictable nature of systems in tran-
sition can initiate a range of emotional responses such as fear, panic, or even hope and

optimism. At the same time, a transition can benefit or harm a system as a significant change does not always lead to positive outcomes. Therefore, the outcome of revolutionary periods may differ depending on the type of system (Gersick, 1991).

The Theory of Punctuated Equilibrium has been adopted in many disciplines besides biology, such as anthropology (Cachel, 1989), management, information systems (Lyytinen and Newman, 2008), and public policy (Eissler et al., 2016; Workman et al., 2009). The Theory of Punctuated Equilibrium has been able to explain numerous phenomena such as the evolution of a bacterial population (Chaudhuri and Bose, 1999), the diffusion of an innovation (Boushey, 2012), and the relationship between systems analysts and users (Newman and Robey, 1992), as well as to explain the patterns of change and stability in international policy-making (Lundgren et al., 2018). In addition, Punctuated Equilibrium Theory has been applied to explore how organisations or industries undergo a sudden rapid change (Sastry, 1997; Street and Meister, 2004)), and how the behavior of project groups rapidly changes across time (Gersick, 1988; Chang et al., 2003). Furthermore, the Theory of Punctuated Equilibrium is invoked to explain online behaviours and phenomena. For example, Jarvenpaa et al. (2004) used the Punctuated Equilibrium Theory to examine how trust affects the attitudes and performance of global virtual teams. In particular, the study theorises that before a specific transition point, trust has a direct effect on the attitudes and the performance of virtual teams. However, after the transition point, trust has a moderating effect on the attitudes and the performance of virtual teams. Other studies have applied the Punctuated Equilibrium Theory to examine how a specific event can trigger the attention to policy issues and push for online mobilisation to support a change in policies or legislations. For instance, Dumas et al. (2015) looked at the Sandy Hook Elementary School shootings as a triggering event to foster online mobilisation that supports renewed attention to gun control legislation. Prior to Sandy Hook, gun control legislation had lain dormant for years in an ongoing state of equilibrium. However, following Sandy Hook, discussions about gun control started to take place in the media to enable people to propose action plans through e-petitioning. Correspondingly, other research studied the UK government petition website and found that a petition might be languished for months and suddenly become popular if a relevant external event triggers it. Thus, the punctuated equilibrium model has been embraced to elucidate the findings, as it implies that policy attention of any issue will be characterised by a long period of stasis during which little change occurs. However, an event or incident can turn all eyes to a policy and cause actions to hit the policy agenda (Margetts et al., 2016).

Appendix C

Controversy

This appendix shows an example of the controversy between editors as highlighted in Chapter 8. This controversy occurred in the talk pages of a cardiovascular disease article:

- *Not sure why much of the content was removed in this edit.[3] (....) Thus restored again. Doc James (talk · contribs · email) (if I write on your page reply on mine) 10:09, 17 April 2013 (UTC)*

- *It would be useful if the user in question would discuss here to get consensus for changes first. Doc James (talk · contribs · email) (if I write on your page reply on mine) 01:28, 19 April 2013 (UTC)*

- *It's quite evident that there is no ref suggesting a controversy. Another ref indicates a problem with replacing saturated fat with refined sugars, which seems to be a no-brainer. That does not eliminate sat fat as problematic though. — Preceding unsigned comment added by Fgmoon353 (talk • contribs) 19:39, 19 April 2013 (UTC)*

- *We have meta analyses that state different things. This one states "that there is no significant evidence for concluding that dietary saturated fat is associated with an increased risk of CHD or CVD"[4] (http://www.ncbi.nlm.nih.gov/pubmed/....) Doc James (talk · contribs · email) (if I write on your page reply on mine) 01:17, 20 April 2013 (UTC)*

- *And then we have this just out "An updated meta-analysis of linoleic acid intervention trials showed no evidence of cardiovascular benefit."[5] (http://www.ncbi .nlm.nih.gov/pubmed/23386268). Doc James (talk · contribs · email) (if I write on your page reply on mine) 01:26, 20 April 2013 (UTC)*

– *Well, the role of saturated fat in CHD is pretty well established (every major medical organization recognizes it as a risk factor for CHD); the more recent research has suggested those who substitute it with carbo-hydrates, however, do not get much (if any) benefits as compared to replacing it with mono or poly unsaturated fat (see for example these recent reviews [6] (http://www.ncbi.nlm. nih.gov/pubmed/....) [7] (http://www.ncbi.nlm.nih.gov/....) [8] (http://www.ncbi.nlm.nih.gov/....). That saturated fat doesn't have any effect on CHD is a WP:FRINGE position, and suggesting there is a significant controversy here probably gives undue weight to that position. Yobol (talk) 01:33, 20 April 2013 (UTC)*

– *Again, these studies suggest replacing saturated fat with refined carbo-hydrates (sugar) as problematic, I mean really? Is that a mystery to anyone? Replacing saturated fat dietary calories with whole-food carbs, like leafy vegetables (also a carb) does not carry the same problematic outcome. I think it's unscientific to throw all carbohydrates in the same bucket, which is what seems be going on here*

– *The above updated meta analysis from 2013 which was published in the BMJ and found no benefit from substituting poly unsaturated for sat-urated fat. The BMJ is not known for publishing fringe opinions. I agree that the Cochrane review from 2012 provides 'suggestive' evidence of benefit and should be included. The review from BMC states "al-though major flaws have been reported in the analyses supporting this approach" Doc James (talk · contribs · email) (if I write on your page reply on mine) 06:18, 20 April 2013 (UTC)*

– *The sentence reflecting the Cocharan meta isn't reflected of his conclu-sions. When conclusions are evaluated, whats the relevance? It seems to further prove saturated fat as problematic. It certainly doesn't back up the preceding sentence well.*

– *And than we have this 2012 JAMA systematic review and meta anal-ysis [9] (http://www.ncbi.nlm.nih.gov/....) which concluded "omega-3 PUFA supplementation was not associated with a lower risk of all-cause mortality, cardiac death, sudden death, myocardial infarction, or stroke based on relative and absolute measures of association." Doc James (talk · contribs ·email) (if I write on your page reply on mine) 06:54, 20 April 2013 (UTC)*

– *Of course, we aren't talking about supplementation of diet with omega 3, we are discussing replacing saturated fats with PUFA. In fact, other reviews have noted replacing SFA with PUFA may be independent of omega 6 [10] (http://www.ncbi.nlm.nih.gov/....). I think we have to be*

careful in how we phrase this section. The "controversy" in the medical community is the type of nutrient to replace SFA with to improve CV risk (and there does not appear to be much controversy here; most recommendations including the AND now suggest replacing with PUFA instead of a straight reduction in SFA). There is also a "controversy" in the lay public where there is a suggestion that SFA plays absolutely no role in CV risk; this is clearly the fringe position we should be clear about not supporting. Yobol (talk) 13:57, 20 April 2013 (UTC)

- *Yes the majority of reviews recommend replacing SFA with PUFA and state that there is tentative evidence that it decreases CVD. We have this review from 2010 "A meta-analysis of prospective epidemiological studies showed that there is no significant evidence for concluding that dietary saturated fat is associated with an increased risk of CHD or CVD" [11] (http://www.ncbi. nlm.nih.gov/....) Not sure how reputable the journal though. Doc James (talk · contribs · email) (if I write on your page reply on mine) 14:34, 20 April 2013 (UTC)*

- *Your first sentence is sound, but this 2010 meta study was conducted by members heavily funded by the meat and dairy industries, and the Atkins foundation. It was found to follow improper procedures, and has been largely discredited. J. Stamler is quite highly respected, and stated his full opposition to it. His findings are linked*

- *The paper in question states "No conflicts of interest were reported." Doc James (talk · contribs · email) (if I write on your page reply on mine) 13:54, 23 April 2013 (UTC)*

- *Keep in mind most of these clinical trials involve tests on the elderly and or, those with prior CVD events, usually both. Also, many of these researchers aren't debunking SFA as a leading risk factor for CVD, but determining what EFA to replace those dietary calories with. What about replacing those calories with complex carbohydrates, in the form of green leafy vegetables, whole fruits, and whole-grains? Also, are there any clinical trials with an ideal control population, perhaps patients with normal/acceptable/borderline blood panels, blood pressure, and weight, perhaps 20-40 years old and observe their findings? Where are the trials which get an otherwise normal patient from an LDL of 150 (acceptable), down to below 75 (outstanding)? This would be the ultimate in preventive medicine, finding a 30 year old male, and getting his numbers down to exceptional levels at 30 years old, instead of waiting for his second heart attack at 65, and trying a litany of measures based on these clinical trials. There are numerous trials and epidemiological research indicating the benefits of lifetime low cholesterol panels vs a lifetime of high panels. A patient with a sudden drop after a CVD event*

scares the patient into eliminating risk factors seems to be the focus in many of these trials. Even still, some CVD patients have greatly bene-fited from an aggressive dietary treatment, through elimination of SFA. While anecdotal, an example being President Bill Clinton, adopting a mostly strict vegetarian diet, which eliminates all SFA. I believe this section would benefit from a rewrite. I suggest breaking this into mul-tiple paragraphs, the first being the reduction of saturated fat, and it's benefits. Para number 2 should show popular diets that achieve desired outcomes through the reduction of SFA. Again, I see no hard reason to suggest referencing the supposed "saturated fat controversy" in any case.

– *I do not know why you 1. Added this primary research paper from 2003 looking at 46 people [12] (http://www.ncbi.nlm.nih.gov /....) I see what you're saying here. I wanted to grab something showing dietary benefit in preventing heart disease. I'll remove this one, and substitute an-other. 1. Removed this 2012 review article from the Journal of Internal Medicine [13] (http://www.ncbi.nlm.nih .gov/....) Was unaware that I did. I can put it back. 1. Also IMO we should discuss overall diets before we discuss specific dietary components. For example? The diet section about med diet et al seems a bit 'thrown together' is that what you mean? So have reverted your changes Doc James (talk · contribs · email) (if I write on your page reply on mine) 03:38, 23 April 2013 (UTC)*

– *Why not just add back the old stuff, and remove the new stuff? I also corrected some flow and structure issues. Which flow and structure issues exactly did you correct? Unsure why you are trying to change the order of the sections? Doc James (talk · contribs · email) (if I write on your page reply on mine) 13:51, 23 April 2013 (UTC)*

– *Refs are needed to support the additions you made including "Often overlooked are the patient's results based off what nutrient replaces di-etary saturated fat." "However, little or no benefit is likely if saturated fat is replaced by carbohydrate, but this will in part depend on the form of carbohydrate. Because both N-6 and N-3 polyunsaturated fatty acids are essential and reduce risk of heart disease, the ratio of N-6 to N-3 is not useful and can be misleading. In practice, reducing red meat and dairy products in a food supply and increasing intakes of nuts, fish, soy products and non-hydrogenated vegetable oils will improve the mix of fatty acids and have a markedly beneficial effect on rates of CVD." Doc James (talk · contribs · email) (if I write on your page reply on mine) 14:08, 23 April 2013 (UTC)*

– *Okay found it. You copied and pasted from [14] (http://www.ncbi.nlm.nih.gov/....).*
  You cannot do this as it is plagiarism. Doc James (talk · contribs ·
  email) (if I write on your page reply on mine) 14:09, 23 April 2013
  (UTC)

– *I copied and pasted the text so I could read it on one screen. I just*
  simply forgot to remove it, but I have now. I never intended to leave it.
  I am aware that C and P isn't allowed. — Preceding unsigned comment
  added by Fgmoon353 (talk • contribs) 14:13, 23 April 2013 (UTC)

– *Your changes do not make sense. Have reported you for edit war-*
  ring here [15] (http://en.wikipedia.org/....)Doc James (talk · contribs
  · email) (if I write on your page reply on mine) 14:34, 23 April 2013
  (UTC)

– *I have reviewed your criticisms, yet you still do not like my edits, be-*
  cause, well apparently, you just don't. That's not good enough, I engaged
  your criticisms (3 questions) in good faith, and made changes. If my
  changes don't make sense, and I've corrected the issues you had, yet you
  still have problems, do you have additional issues, if so what are they?
  I welcome Wikipedia's recommendation. I've also asked you a litany of
  questions, to which you might have answered one or two. It sounds like
  you are having control issues because you aren't getting your way, I'm
  sorry about that. To put this more simply, you're not the gate keeper to
  this page. — Preceding unsigned comment added by Fgmoon353 (talk
  •contribs) 14:42, 23 April 2013 (UTC)

– *I have reverted the latest change (http://en.wikipedia.org/....) that re-*
  moved well-sourced content and changed existing content to move away
  from what sources say. In particular, important content about the unset-
  tled question over the effects of saturated fat was made less clear, some
  unnecessary editorializing was added ("Often overlooked..."), changes
  against WP:MEDMOS were made (references to "patient"), the JAMA
  meta-analysis was described as focusing on elderly women but it appears
  the study base was wider than that, and the mention "Total fat intake
  does not appear to be an important risk factor" was removed. Suggest
  we break up and discuss this large number of significant changes to get
  consensus so the content can be stable. Zad68 15:07, 23 April 2013
  (UTC)

– *Unsettled question over the effects of saturated fat was made less clear*
  In what respect? The is no question on the effects of replacing saturated
  fat with sugary junk food, and that is well sited. Any notion that sug-
  gests that saturated fat is not a leading cause of CVD is WP:FRINGE
  opinion at best. Not to mention that it contradicts several statements
  on this page which repeatedly state that it is. JAMA meta-analysis was

described as focusing on elderly women but it appears the study base was wider than that Admitted this, and changed "Total fat intake does not appear to be an important risk factor" This statement vastly understates the ref. It's misleading and suggests that dietary fat is not potentially problematic. "total fat intake" means "total fat intake". Our readers are intelligent and can figure this out. Especially when the next few sentences adds details. Doc James (talk · contribs · email) (if I write on your page reply on mine) 16:05, 23 April 2013 (UTC)

– *("Often overlooked...") How does this violate wiki standards? — Preceding unsigned comment added by Fgmoon353 (talk ● contribs) 15:25, 23 April 2013 (UTC) Were is the evidence that it is "often overlooked"? If there is not any it is original research. Doc James (talk · contribs · email) (if I write on your page reply on mine) 16:06, 23 April 2013 (UTC)*

– *"There however is some questions around the effect of saturated fat on cardiovascular disease in the medical literature." This emphasizes WP:FRINGE opinions on saturated fat, directing to saturated fat controversy. The word 'some questions' leads to believe it's substantial enough to be noticed in medical literature. It's absolutely miniscule at best. Nutritional science has moved on from this debate. Your edits are a little hard to follow because you not only change a bunch of text but move it around in a single edit. IMO we should discuss overall diets first and than specific components second. You keep switching it around and have not provided any justification or developed consensus. Doc James (talk · contribs · email) (if I write on your page reply on mine) 16:05, 23 April 2013 (UTC)*

– *Can you not address my statement? — Preceding unsigned comment added by Fgmoon353 (talk ● contribs) 03:35, 28 April 2013 (UTC)*

Appendix D

Uncertainty and Controversy

This appendix explores the relationship between uncertainty and controversy on Wikipedia:

Controversies are simply defined as "situations where actors disagree" (Venturini, 2010). Controversies begin when actors are unable to ignore each other and end when actors discover a common ground and reach an agreement. Anything between three two points can be called a controversy (Venturini, 2010). Controversies among editors usually take place on Wikipedia talk pages where editors are given the chance to discuss and resolve an issue related to editing articles. The following section examines the controversies on Wikipedia talk pages in relationship to different types of uncertainty. This firstly entails operationalising the concept of controversies in order to code instances of controversies derived directly from the text. Therefore, controversies are identified as any discussion on Wikipedia talk pages that includes a disagreement between more than two editors. In particular, controversies are any discussions that range from a simple difference in opinion to a prolonged dispute or heated discussion. However, controversies here are not simply arguments between two editors (d'Alembert, 2015). In fact, controversies are only considered to be such if they involve at least three different editors as identified from their distinct usernames. Appendix C shows an example of a controversy between editors in a Wikipedia talk page of the cardiovascular disease article: Table D.1 looks at the number of posts that include a controversy between editors in new disease outbreak articles in comparison to chronic diseases articles. The table indicates that the talk pages of chronic diseases articles contain more controversial posts than the talk pages of new disease outbreak articles.

After identifying posts that include a controversy between editors, these posts have also been examined to check whether they include forms of uncertainty identified as scientific uncertainty, references uncertainty, and conflicting references. Table D.2 depicts the contingency table to show the results of the intersections between a controversy and an uncertainty in different posts on the Wikipedia talk pages. The results suggest that editors can experience an uncertainty, which can lead to a controversy between them. The example on Appendix C illustrates that editors might experience conflicting references uncertainty as a result of finding publications that can state contradictory information about the association between dietary saturated fat and the increased risk of CHD or CVD. This case of uncertainty also provokes a controversy between editors.

Articles	**Controversies**
New disease outbreaks	3
Chronic diseases	49

Table D.1: Number of Posts that Contain a Controversy between Editors in New Disease Outbreaks and Chronic Diseases articles

	Controversies	**Non-controversies**
Uncertainty	26	75
Non-uncertainty	26	1559

Table D.2: Contingency Table of the Frequency of Controversy and Uncertainty in New Disease Outbreaks and Chronic Diseases articles

Bibliography

B. Thomas Adler, Krishnendu Chatterjee, Luca de Alfaro, Marco Faella, Ian Pye, and Vishwanath Raman. Assigning Trust to Wikipedia Content. In *Proceedings of the 4th International Symposium on Wikis*, WikiSym '08, pages 26:1–26:12, New York, NY, USA, 2008. ACM. ISBN 978-1-60558-128-6.

B. Thomas Adler and Luca de Alfaro. A Content-driven Reputation System for the Wikipedia. In *Proceedings of the 16th International Conference on World Wide Web*, WWW '07, pages 261–270, New York, NY, USA, 2007. ACM. ISBN 978-1-59593-654-7.

V. Ahmed, A. Opoku, and Z. U. H. Aziz. *Research Methodology in the Built Environment: A Selection of Case Studies*. Routledge - Taylor and Francis Group, UK, US, March 2016. ISBN 978-1-317-53425-9.

Byung Gyu Ahn, Benjamin Van Durme, and Chris Callison-Burch. WikiTopics: What is Popular on Wikipedia and Why. In *Proceedings of the Workshop on Automatic Summarization for Different Genres, Media, and Languages*, WASDGML '11, pages 33–40, Stroudsburg, PA, USA, 2011. Association for Computational Linguistics. ISBN 978-1-932432-94-7.

Mohammad Akbarpour and Matthew O. Jackson. Diffusion in Networks and the Virtue of Burstiness. *Proceedings of the National Academy of Sciences*, 115(30):E6996–E7004, July 2018. ISSN 0027-8424, 1091-6490.

B. Akhgar, D. Fortune, R. E. Hayes, B. Guerra, and M. Manso. 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)*, pages 760–765, November 2013.

Reham Al Tamime, Richard Giordano, and Wendy Hall. Humans and Bots in Controversial Environments: A Closer Look at Their Interactions. *Qatar Foundation Annual Research Conference Proceedings*, 2018(3):ICTPD425, March 2018a.

Reham Al Tamime, Richard Giordano, and Wendy Hall. Observing Burstiness in Wikipedia Articles during New Disease Outbreaks. In *Proceedings of the 10th ACM Conference on Web Science - WebSci '18*, pages 117–126, Amsterdam, Netherlands, 2018b. ACM Press. ISBN 978-1-4503-5563-6.

Reham Al Tamime, Richard Giordano, and Wendy Hall. Uncertainty During New Disease Outbreaks in Wikipedia. *Proceedings of the International AAAI Conference on Web and Social Media*, 13(01):38–46, June 2019. ISSN 2334-0770.

Michael J. Albers. Communication As Reducing Uncertainty. In *Proceedings of the 30th ACM International Conference on Design of Communication*, SIGDOC '12, pages 1–8, New York, NY, USA, 2012. ACM. ISBN 978-1-4503-1497-8.

Gary Albrecht, Ray Fitzpatrick, and Susan Scrimshaw. *Handbook of Social Studies in Health and Medicine*. London, 2000.

Rodrigo B Almeida, Barzan Mozafari, and Junghoo Cho. On the Evolution of Wikipedia. In *Proceedings of ICWSM'07*, March 2007.

Tim Althoff, Damian Borth, Jörn Hees, and Andreas Dengel. Analysis and Forecasting of Trending Topics in Online Media Streams. In *Proceedings of the 21st ACM International Conference on Multimedia*, MM '13, pages 907–916, New York, NY, USA, 2013. ACM. ISBN 978-1-4503-2404-5.

Dilanthi Amaratunga, D. Baldry, M. Sarshar, and R. A. Newton. Qualitative and Quantitative Research in the Built Environment: Application of "Mixed" Research Approach: A Conceptual Framework to Measure FM Performance. *Work Study*, 20: 17–31, January 2002. ISSN 0043-8022.

Judd Antin. My Kind of People?: Perceptions About Wikipedia Contributors and Their Motivations. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '11, pages 3411–3420, New York, NY, USA, 2011. ACM. ISBN 978-1-4503-0228-9.

Judd Antin, Raymond Yee, Coye Cheshire, and Oded Nov. Gender Differences in Wikipedia Editing. In *Proceedings of the 7th International Symposium on Wikis and Open Collaboration*, WikiSym '11, pages 11–14, New York, NY, USA, 2011. ACM. ISBN 978-1-4503-0909-7.

Patrik Aspers. Forms of Uncertainty Reduction: Decision, Valuation, and Contest. *Theory and Society*, 47(2):133–149, April 2018. ISSN 1573-7853.

Juozas Auskalnis, Nerijus Paulauskas, and Algirdas Baskys. Application of Local Outlier Factor Algorithm to Detect Anomalies in Computer Network. *Elektronika ir Elektrotechnika*, 24(3):96–99–99, June 2018. ISSN 2029-5731.

Lucinda Austin, Brooke Fisher Liu, and Yan Jin. How Audiences Seek Out Crisis Information: Exploring the Social-Mediated Crisis Communication Model. *Journal of Applied Communication Research*, 40(2):188–207, May 2012. ISSN 0090-9882, 1479-5752.

Samy A. Azer, Nourah M. AlSwaidan, Lama A. Alshwairikh, and Jumana M. AlShammari. Accuracy and Readability of Cardiovascular Entries on Wikipedia: Are They Reliable Learning Resources for Medical Students? *BMJ Open*, 5(10):e008187, October 2015. ISSN 2044-6055, 2044-6055.

Austin S. Babrow. Problematic Integration Theory. In *The International Encyclopedia of Interpersonal Communication*, pages 1–9. American Cancer Society, 2015. ISBN 978-1-118-54019-0.

Per Bak and Stefan Boettcher. Self-Organized Criticality and Punctuated Equilibria. *Physica D: Nonlinear Phenomena*, 107(2-4):143–150, September 1997. ISSN 01672789. arXiv: cond-mat/9701157.

Martina Balestra, Coye Cheshire, Ofer Arazy, and Oded Nov. Investigating the Motivational Paths of Peer Production Newcomers. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, CHI '17, pages 6381–6385, New York, NY, USA, 2017. ACM. ISBN 978-1-4503-4655-9.

Irad Ben-Gal. Outlier Detection. In Oded Maimon and Lior Rokach, editors, *Data Mining and Knowledge Discovery Handbook*, pages 131–146. Springer US, Boston, MA, 2005. ISBN 978-0-387-25465-4.

Yakov Ben-Haim. Uncertainty, Probability and Information-Gaps. *Reliability Engineering & System Safety*, 85(1):249–266, July 2004. ISSN 0951-8320.

Arcelio Benetoli, Timothy Frank Chen, Marion Schaefer, Betty B Chaar, and Parisa Aslani. Professional Use of Social Media by Pharmacists: A Qualitative Study. *Journal of Medical Internet Research*, 18(9), September 2016. ISSN 1439-4456.

Charles R. Berger and Richard J. Calabrese. Some Explorations in Initial Interaction and Beyond: Toward a Developmental Theory of Interpersonal Communication. *Human Communication Research*, 1(2):99–112, December 1975. ISSN 1468-2958.

Tim Berners-Lee and Mark Fischetti. *Weaving the Web; The Original Design and Ultimate Destiny of the World Wide Web by Its Inventor (2 Cassettes)*. Harper Audio, 1999. ISBN 978-0-694-52125-8.

David M. Berry, editor. *Understanding Digital Humanities*. Palgrave Macmillan UK, London, 2012. ISBN 978-0-230-29265-9 978-0-230-37193-4.

T. C. Biggs, N. Jayakody, K. Best, and E. V. King. Quality of Online Otolaryngology Health Information. *The Journal of Laryngology and Otology*, 132(6):560–563, June 2018. ISSN 1748-5460.

Joshua E. Blumenstock. Size Matters: Word Count As a Measure of Quality on Wikipedia. In *Proceedings of the 17th International Conference on World Wide Web*, WWW '08, pages 1095–1096, New York, NY, USA, 2008. ACM. ISBN 978-1-60558-085-2.

P. Bogdanov, N. D. Larusso, and A. Singh. Towards Community Discovery in Signed Collaborative Interaction Networks. In *2010 IEEE International Conference on Data Mining Workshops*, pages 288–295, December 2010.

Erik Borra, Esther Weltevrede, Paolo Ciuccarelli, Andreas Kaltenbrunner, David Laniado, Giovanni Magni, Michele Mauri, Richard Rogers, and Tommaso Venturini. Societal Controversies in Wikipedia Articles. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*, CHI '15, pages 193–196, New York, NY, USA, 2015. ACM. ISBN 978-1-4503-3145-6.

Graeme Boushey. Punctuated Equilibrium Theory and the Diffusion of Innovations. *Policy Studies Journal*, 40(1):127–146, February 2012. ISSN 1541-0072.

George Edward Pelham Box and Gwilym Jenkins. *Time Series Analysis, Forecasting and Control*. Holden-Day, Inc., San Francisco, CA, USA, 1990. ISBN 978-0-8162-1104-3.

James J. Bradac. Theory Comparison: Uncertainty Reduction, Problematic Integration, Uncertainty Management, and Other Curious Constructs. *Journal of Communication*, 51(3):456–476, September 2001. ISSN 1460-2466.

Ulrik Brandes and Jürgen Lerner. Visual Analysis of Controversy in User-Generated Encyclopedias. *Information Visualization*, 7(1):34–48, March 2008. ISSN 1473-8716.

Dale E. Brashers. Communication and Uncertainty Management. *Journal of Communication*, 51(3):477–497, September 2001. ISSN 1460-2466.

Dale E. Brashers and Timothy P. Hogan. The Appraisal and Management of Uncertainty: Implications for Information-Retrieval Systems. *Information Processing & Management*, 49(6):1241–1249, November 2013. ISSN 0306-4573.

Dale E. Brashers, Judith L. Neidig, Stephen M. Haas, Linda K. Dobbs, Linda W. Cardillo, and Jane A. Russell. Communication in the Management of Uncertainty: The Case of Persons Living with HIV or AIDS. *Communication Monographs*, 67(1):63–84, March 2000. ISSN 0363-7751.

Francesco Brigo and Roberto Erro. Why Do People Google Movement Disorders? An Infodemiological Study of Information Seeking Behaviors. *Neurological Sciences: Official Journal of the Italian Neurological Society and of the Italian Society of Clinical Neurophysiology*, 37(5):781–787, May 2016. ISSN 1590-3478.

Francesco Brigo, Stanley C. Igwe, Raffaele Nardone, Piergiorgio Lochner, Frediano Tezzon, and Willem M. Otte. Wikipedia and Neurological Disorders. *Journal of Clinical Neuroscience*, 22(7):1170–1172, July 2015a. ISSN 0967-5868, 1532-2653.

Francesco Brigo, Willem M. Otte, Stanley C. Igwe, Harald Ausserer, Raffaele Nardone, Frediano Tezzon, and Eugen Trinka. Information-Seeking Behaviour for Epilepsy: An Infodemiological Study of Searches for Wikipedia Articles. *Epileptic Disorders: International Epilepsy Journal with Videotape*, 17(4):460–466, December 2015b. ISSN 1950-6945.

H. I. L. Brink. Validity and Reliability in Qualitative Research. *Curationis*, 16(2):35–38, March 1993. ISSN 2223-6279.

Paula Brough. *Advanced Research Methods for Applied Psychology : Design, Analysis and Reporting*. Routledge, August 2018. ISBN 978-1-315-51797-1.

Jason Brownlee. How to Create an ARIMA Model for Time Series Forecasting in Python, January 2017.

Michael K. Buckland. Information as Thing. *Journal of the American Society for Information Science*, 42(5):351–360, June 1991. ISSN 1097-4571.

S. Bykau, F. Korn, D. Srivastava, and Y. Velegrakis. Fine-grained Controversy Detection in Wikipedia. In *2015 IEEE 31st International Conference on Data Engineering*, pages 1573–1584, April 2015.

Susan Cachel. The Theory of Punctuated Equilibria and Evolutionary Anthropology. *Journal of Social and Biological Structures*, 12(2):225–239, April 1989. ISSN 0140-1750.

Danielle M. Candelario, Victoria Vazquez, William Jackson, and Timothy Reilly. Completeness, Accuracy, and Readability of Wikipedia as a Reference for Patient Medication Information. *Journal of the American Pharmacists Association*, 57(2):197–200.e1, March 2017. ISSN 1544-3191.

Ashley Castleberry and Amanda Nolen. Thematic Analysis of Qualitative Research Data: Is It as Easy as It Sounds? *Currents in Pharmacy Teaching and Learning*, 10 (6):807–815, June 2018. ISSN 1877-1297.

Boris Cergol and Matjaž Omladič. What can Wikipedia and Google Tell us about Stock Prices under Different Market Regimes? *ARS MATHEMATICA CONTEMPORANEA*, 9(2):301–320, June 2015. ISSN 1855-3974.

Artemis Chang, Prashant Bordia, and Julie Duck. Punctuated Equilibrium and Linear Progression: Toward a New Understanding of Group Development. *The Academy of Management Journal*, 46(1):106–117, 2003. ISSN 0001-4273.

Indranath Chaudhuri and Indrani Bose. Punctuated equilibrium in an evolving bacterial population. *Physica A: Statistical Mechanics and its Applications*, 270(1):63–68, August 1999. ISSN 0378-4371.

Yanguang Chen. Urban Chaos and Replacement Dynamics in Nature and Society. *Physica A: Statistical Mechanics and its Applications*, 402:266–282, May 2014. ISSN 03784371. arXiv: 1106.0887.

F. Chevalier, S. Huot, and J. Fekete. WikipediaViz: Conveying article quality for casual Wikipedia readers. In *2010 IEEE Pacific Visualization Symposium (PacificVis)*, pages 49–56, March 2010.

Ji Young Cho and Eun-Hee Lee. Reducing Confusion about Grounded Theory and Qualitative Content Analysis: Similarities and Differences. *The Qualitative Report*, 19(32):1–20, 2014.

Marek Ciglan and Kjetil Nirvåg. WikiPop: Personalized Event Detection System Based on Wikipedia Page View Statistics. In *Proceedings of the 19th ACM International Conference on Information and Knowledge Management*, CIKM '10, pages 1931–1932, New York, NY, USA, 2010. ACM. ISBN 978-1-4503-0099-5.

Kevin A Clauson, Hyla H Polen, Maged N Kamel Boulos, and Joan H Dzenowagis. Scope, Completeness, and Accuracy of Drug Information in Wikipedia. *Annals of Pharmacotherapy*, 42(12):1814–1821, December 2008. ISSN 1060-0280.

CMU Statistics. lab-05 36-350.pdf - Lab 5 Testing Our Way to Outliers 36-350 Statistical Computing Friday 30 September 2011 Agenda Debugging and Testing Finding, 2013.

Charles Cole. Shannon Revisited: Information in Terms of Uncertainty. *Journal of the American Society for Information Science*, 44(4):204–211, May 1993. ISSN 1097-4571.

Ulrike Cress and Joachim Kimmerle. A systemic and Cognitive View on Collaborative Knowledge Building with Wikis. *International Journal of Computer-Supported Collaborative Learning*, 3(2):105, January 2008. ISSN 1556-1615.

John Creswell, Vickie L Plano Clark, Michelle Gutmann, and William E Hanson. *Advanced Mixed Methods Research Designs*, pages 209–240. 01 2003. ISBN 9780761920731.

John W. Creswell. *Qualitative Inquiry & Research Design: Choosing among Five Approaches*. Sage Publications, Thousand Oaks, 2nd ed. edition, 2007. ISBN 978-1-4129-1606-6.

John W Creswell. Controversies in Mixed Methods Research. *The Sage Handbook of Qualitative Research*, 4:269–284, 2011.

John W. Creswell. *Research Design: Qualitative, Quantitative, and Mixed Methods Approaches*. SAGE Publications, 2014. ISBN 978-1-4522-2609-5.

John W. Creswell and Vicki L. Plano Clark. *Designing and Conducting Mixed Methods Research*. Designing and Conducting Mixed Methods Research. Sage Publications, Inc, Thousand Oaks, CA, US, 2007. ISBN 978-1-4129-2792-5 978-1-4129-2791-8.

John W. Creswell and Vicki L. Plano Clark. *Designing and Conducting Mixed Methods Research*. SAGE Publications, Los Angeles, 2011. ISBN 978-1-4129-7517-9 1-4129-7517-4.

Tom Cross. Puppy Smoothies: Improving the Reliability of Open, Collaborative Wikis. *First Monday*, 11(9), September 2006. ISSN 13960466.

Jianhua Dai and Qing Xu. Approximations and Uncertainty Measures in Incomplete Information Systems. *Information Sciences*, 198:62–80, September 2012. ISSN 0020-0255.

Jean-Baptiste le Rond d'Alembert. Dispute, Argument, Debate, Altercation. *Encyclopedia of Diderot & d'Alembert - Collaborative Translation Project*, January 2015.

Kajsa E. Dalrymple, Rachel Young, and Melissa Tully. "Facts, Not Fear": Negotiating Uncertainty on Social Media During the 2014 Ebola Crisis. *Science Communication*, 38(4):442–467, August 2016. ISSN 1075-5470, 1552-8545.

Charles Darwin. *On the Origin of Species by Means of Natural Selection, or, The Preservation of Favoured Races in the Struggle for Life*. John Murray, London:, 1859.

Sanmay Das, Allen Lavoie, and Malik Magdon-Ismail. Manipulation Among the Arbiters of Collective Intelligence: How Wikipedia Administrators Mold Public Opinion. *ACM Trans. Web*, 10(4):24:1–24:25, December 2016. ISSN 1559-1131.

David De Roure, Clare Hooper, Kevin Page, S©gol®ne Tarte, and Pip Willcox. Observing Social Machines Part 2: How to Observe? In *Proceedings of the ACM Web Science Conference*, WebSci '15, pages 13:1–13:5, New York, NY, USA, 2015. ACM. ISBN 978-1-4503-3672-7. event-place: Oxford, United Kingdom.

Brian de Silva and Ryan Compton. Prediction of Foreign Box Office Revenues Based on Wikipedia Page Activity. *arXiv:1405.5924 [physics]*, May 2014. arXiv: 1405.5924.

Norman K. Denzin and Yvonna S. Lincoln. *Strategies of Qualitative Inquiry*. SAGE Publications, February 2003. ISBN 978-0-7619-2691-7. Google-Books-ID: vRZHAAAA-MAAJ.

Cambridge Dictionary. Burst Meaning in the Cambridge English Dictionary. `https://dictionary.cambridge.org/dictionary/english/burst`, 2018. Last accessed 25 November 2018.

Pierpaolo Dondio, Stephen Barrett, Stefan Weber, and Jean Marc Seigneur. Extracting Trust from Domain Analysis: A Case Study on the Wikipedia Project. In Laurence T. Yang, Hai Jin, Jianhua Ma, and Theo Ungerer, editors, *Autonomic and Trusted Computing*, Lecture Notes in Computer Science, pages 362–373. Springer Berlin Heidelberg, 2006. ISBN 978-3-540-38622-3.

Larry M. Dooley. Case Study Research and Theory Building. *Advances in Developing Human Resources*, 4(3):335–354, August 2002. ISSN 1523-4223.

Louise Doyle, Anne-Marie Brady, and Gobnait Byrne. An Overview of Mixed Methods Research. *Journal of Research in Nursing*, March 2009.

Julie Dugdale, Bartel Van de Walle, and Corinna Koeppinghoff. Social Media and SMS in the Haiti Earthquake. In *Proceedings of the 21st International Conference on World Wide Web*, WWW '12 Companion, pages 713–714, New York, NY, USA, 2012. ACM. ISBN 978-1-4503-1230-1.

Duke University. Introduction to ARIMA: Nonseasonal Models, 2019.

Catherine Dumas, Daniel LaManna, Teresa Harrison, S. Ravi, Loni Hagen, Christopher Kotfila, and Feng Chen. E-petitioning as Collective Political Action in We the People. *iConference 2015 Proceedings*, pages 1–20, January 2015.

Kimberly J. Dunham-Snary, Michael W. Sandel, David G. Westbrook, and Scott W. Ballinger. A Method for Assessing Mitochondrial Bioenergetics in Whole White Adipose Tissues. *Redox Biology*, 2:656–660, January 2014. ISSN 2213-2317.

F. Dzogang, T. Lansdall-Welfare, and N. Cristianini. Seasonal Fluctuations in Collective Mood Revealed by Wikipedia Searches and Twitter Posts. In *2016 IEEE 16th International Conference on Data Mining Workshops (ICDMW)*, pages 931–937, December 2016.

Sara E Shaw and Julia Bailey. Discourse Analysis: What Is It and Why Is It Relevant to Family Practice? *Family Practice*, 26(5):413–419, October 2009. ISSN 0263-2136.

Rebecca Eissler, Annelise Russell, and Bryan D. Jones. The Transformation of Ideas: The Origin and Evolution of Punctuated Equilibrium Theory. In B. Guy Peters and Philippe Zittoun, editors, *Contemporary Approaches to Public Policy: Theories, Controversies and Perspectives*, International Series on Public Policy, pages 95–112. Palgrave Macmillan UK, London, 2016. ISBN 978-1-137-50494-4.

Michael D. Ekstrand and John T. Riedl. Rv You're Dumb: Identifying Discarded Work in Wiki Article History. In *Proceedings of the 5th International Symposium on Wikis and Open Collaboration*, WikiSym '09, pages 4:1–4:10, New York, NY, USA, 2009. ACM. ISBN 978-1-60558-730-1.

Niles Eldredge. The Early "Evolution" of "Punctuated Equilibria". *Evolution: Education and Outreach*, 1(2):107–113, April 2008. ISSN 1936-6434.

Niles Eldredge and Stephen Gould. Punctuated Equilibria: An Alternative to Phyletic Gradualism. In *Models in Paleobiology*, pages 82–115. Freeman, Cooper, & Co., San Francisco CA, 1972.

Kerstin Eriksson and Allan McConnell. Contingency Planning for Crisis Management: Recipe for Success or Political Fantasy? *Policy and Society*, 30(2):89–99, May 2011. ISSN 1449-4035, 1839-3373.

Don Fallis. Toward an Epistemology of Wikipedia. *Journal of the American Society for Information Science and Technology*, 59(10):1662–1674, August 2008. ISSN 1532-2890.

A. Fariza, J. A. N. Hasim, and H. Rahadianto. Flood Risk Mapping of Bengawan Solo River in East Java using the Analytic Hierarchy Process and Head/Tails Break Classification. In *2015 International Conference on Science in Information Technology (ICSITech)*, pages 331–336, October 2015.

Michela Ferron and Paolo Massa. Psychological Processes Underlying Wikipedia Representations of Natural and Manmade Disasters. In *Proceedings of the Eighth Annual International Symposium on Wikis and Open Collaboration*, WikiSym '12, pages 2:1–2:10, New York, NY, USA, 2012. ACM. ISBN 978-1-4503-1605-7.

Kathryn C. Finch, Kassandra R. Snook, Carmen H. Duke, King-Wa Fu, Zion Tsz Ho Tse, Atin Adhikari, and Isaac Chun-Hai Fung. Public Health Implications of Social Media Use during Natural Disasters, Environmental Disasters, and other Environmental Concerns. *Natural Hazards*, 83(1):729–760, August 2016. ISSN 0921-030X, 1573-0840.

Carla L. Fisher, Thomas Roccotagliata, Camella J. Rising, David W. Kissane, Emily A. Glogowski, and Carma L. Bylund. "I Don't Want to Be an Ostrich": Managing Mothers' Uncertainty during BRCA1/2 Genetic Counseling. *Journal of Genetic Counseling*, 26(3):455–468, June 2017. ISSN 1573-3599.

Fabian Flöck and Maribel Acosta. WhoVIS: Visualizing Editor Interactions and Dynamics in Collaborative Writing Over Time. In *Proceedings of the 24th International Conference on World Wide Web*, WWW '15 Companion, pages 191–194, New York, NY, USA, 2015. ACM. ISBN 978-1-4503-3473-0.

Heather Ford, Shilad Sen, David R. Musicant, and Nathaniel Miller. Getting to the Source: Where Does Wikipedia Get Its Information From? In *Proceedings of the 9th International Symposium on Open Collaboration*, WikiSym '13, pages 9:1–9:10, New York, NY, USA, 2013. ACM. ISBN 978-1-4503-1852-5.

Gabriel Pui Cheong Fung, Jeffrey Xu Yu, Philip S. Yu, and Hongjun Lu. Parameter Free Bursty Events Detection in Text Streams. In *Proceedings of the 31st International Conference on Very Large Data Bases*, VLDB '05, pages 181–192, Trondheim, Norway, 2005. VLDB Endowment. ISBN 978-1-59593-154-2.

José Galindo. Introduction and Trends to Fuzzy Logic and Fuzzy Databases. *Handbook of Research on Fuzzy Information Processing in Databases*, pages 1–33, 2008.

Peichao Gao, Zhao Liu, Gang Liu, Hongrui Zhao, and Xiaoxiao Xie. Unified Metrics for Characterizing the Fractal Nature of Geographic Features. *Annals of the American Association of Geographers*, 107(6):1315–1331, November 2017. ISSN 2469-4452.

Peichao Gao, Zhao Liu, Meihui Xie, Kun Tian, and Gang Liu. CRG Index: A More Sensitive Ht-Index for Enabling Dynamic Views of Geographic Features. *The Professional Geographer*, 68(4):533–545, October 2016. ISSN 0033-0124.

Claudia Garcia Serpa Osorio-de Castro, Elaine Silva Miranda, Carlos Machado de Freitas, Kenneth Rochel de Camargo, and Hilarie Hartel Cranmer. The Zika Virus Outbreak in Brazil: Knowledge Gaps and Challenges for Risk Reduction. *American Journal of Public Health*, 107(6):960–965, June 2017. ISSN 0090-0036, 1541-0048.

R. Stuart Geiger. The Social Roles of Bots and Assisted Editing Programs. In *Proceedings of the 5th International Symposium on Wikis and Open Collaboration*, WikiSym '09, pages 30:1–30:2, New York, NY, USA, 2009. ACM. ISBN 978-1-60558-730-1.

R. Stuart Geiger and Aaron Halfaker. When the Levee Breaks: Without Bots, What Happens to Wikipedia's Quality Control Processes? In *Proceedings of the 9th International Symposium on Open Collaboration - WikiSym '13*, pages 1–6, Hong Kong, China, 2013. ACM Press. ISBN 978-1-4503-1852-5.

R. Stuart Geiger and David Ribes. The Work of Sustaining Order in Wikipedia: The Banning of a Vandal. In *Proceedings of the 2010 ACM Conference on Computer Supported Cooperative Work*, CSCW '10, pages 117–126, New York, NY, USA, 2010. ACM. ISBN 978-1-60558-795-0.

Nicholas Generous, Geoffrey Fairchild, Alina Deshpande, Sara Y. Del Valle, and Reid Priedhorsky. Global Disease Monitoring and Forecasting with Wikipedia. *PLoS Computational Biology*, 10(11):e1003892, November 2014. ISSN 1553-7358.

Mihai Georgescu, Nattiya Kanhabua, Daniel Krause, Wolfgang Nejdl, and Stefan Siersdorfer. Extracting Event-Related Information from Article Updates in Wikipedia. In Pavel Serdyukov, Pavel Braslavski, Sergei O. Kuznetsov, Jaap Kamps, Stefan Rᵒger, Eugene Agichtein, Ilya Segalovich, and Emine Yilmaz, editors, *Advances in Information Retrieval*, Lecture Notes in Computer Science, pages 254–266. Springer Berlin Heidelberg, 2013. ISBN 978-3-642-36973-5.

John Gerring. *Case Study Research: Principles and Practices*. Cambridge University Press, New York, 2007. ISBN 978-0-521-85928-8.

Connie J. Gersick. Time and Transition in Work Teams: Toward a New Model of Group Development. *Academy of Management Journal*, 31(1):9–41, 1988. ISSN 1948-0989(Electronic),0001-4273(Print).

Connie J. G. Gersick. Revolutionary Change Theories: A Multilevel Exploration of the Punctuated Equilibrium Paradigm. *The Academy of Management Review*, 16(1): 10–36, 1991. ISSN 0363-7425.

Jim Giles. Internet Encyclopaedias Go Head to Head. *Nature*, 438(7070):900–901, December 2005. ISSN 1476-4687.

Lisa Given. *The SAGE Encyclopedia of Qualitative Research Methods*. Thousand Oaks, California, 2008.

Kwang-Il Goh and Albert-Laszlo Barabasi. Burstiness and Memory in Complex Systems. *arXiv:physics/0610233*, October 2006. arXiv: physics/0610233.

Nahid Golafshani. Understanding Reliability and Validity in Qualitative Research. *The Qualitative Report*, 8(4):597–606, December 2003. ISSN 1052-0147.

Niles Eldredge-Stephen Jay Gould and Niles Eldredge. Punctuated Equilibria: An Alternative to Phyletic Gradualism. 1972.

David E. Gray. *Doing Research in the Real World*. SAGE, London, 2014. ISBN 978-1-4462-6018-0 1-4462-6018-6 1-4462-6019-4 978-1-4462-6019-7 978-1-4462-9531-1 1-4462-9531-1.

Xinning Gui, Yubo Kou, Kathleen Pine, Elisa Ladaw, Harold Kim, Eli Suzuki-Gill, and Yunan Chen. 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, pages 214:1–214:14, New York, NY, USA, 2018. ACM. ISBN 978-1-4503-5620-6.

Xinning Gui, Yubo Kou, Kathleen H. Pine, and Yunan Chen. 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*, pages 4520–4533, Denver, Colorado, USA, 2017. ACM Press. ISBN 978-1-4503-4655-9.

Hanning Guo, Scott Weingart, and Katy Börner. Mixed-indicators model for identifying emerging research areas. *Scientometrics*, 89(1):421–435, October 2011. ISSN 1588-2861.

Jheser Guzman and Barbara Poblete. On-line Relevant Anomaly Detection in the Twitter Stream: An Efficient Bursty Keyword Detection Model. In *Proceedings of the ACM SIGKDD Workshop on Outlier Detection and Description*, ODD '13, pages 31–39, New York, NY, USA, 2013. ACM. ISBN 978-1-4503-2335-2.

Carol A. Haigh. Wikipedia as an Evidence Source for Nursing and Healthcare Students. *Nurse Education Today*, 31(2):135–139, February 2011. ISSN 1532-2793.

Alexander Halavais and Derek Lackaff. An Analysis of Topical Coverage of Wikipedia. *Journal of Computer-Mediated Communication*, 13(2):429–440, 2008.

Aaron Halfaker, R. Stuart Geiger, Jonathan T. Morgan, and John Riedl. The Rise and Decline of an Open Collaboration System: How Wikipedia's Reaction to Popularity Is Causing Its Decline. *American Behavioral Scientist*, 57(5):664–688, May 2013. ISSN 0002-7642.

Michael Harwell. Research Design in Qualitative/Quantitative/Mixed Methods. In *The SAGE Handbook for Research in Education: Pursuing Ideas as the Keystone of Exemplary Inquiry*, pages 147–164. SAGE Publications, Inc., Thousand Oaks, 2 edition, 2011.

James M Heilman, Eckhard Kemmann, Michael Bonert, Anwesh Chatterjee, Brent Ragar, Graham M Beards, David J Iberri, Matthew Harvey, Brendan Thomas, Wouter Stomp, Michael F Martone, Daniel J Lodge, Andrea Vondracek, Jacob F de Wolff, Casimir Liber, Samir C Grover, Tim J Vickers, Bertalan Meskó, and Michaël R Laurent. Wikipedia: A Key Tool for Global Public Health Promotion. *Journal of Medical Internet Research*, 13(1), January 2011. ISSN 1438-8871.

James M. Heilman and Andrew G. West. Wikipedia and Medicine: Quantifying Readership, Editors, and the Significance of Natural Language. *Journal of Medical Internet Research*, 17(3):e62, March 2015. ISSN 1438-8871.

Gregor Heinrich. Parameter estimation for text analysis. Technical report, Technical report, 2005.

Jon C. Helton, Jay D. Johnson, William L. Oberkampf, and Cédric J. Sallaberry. Representation of Analysis Results Involving Aleatory and Epistemic Uncertainty. *International Journal of General Systems*, 39(6):605–646, August 2010. ISSN 0308-1079.

Sharlene Nagy Hesse-Biber. *Mixed Methods Research: Merging Theory with Practice*. Guilford Press, 2010.

Thomas Heverin and Lisl Zach. *Microblogging for Crisis Communication: Examination of Twitter Use in Response to a 2009 Violent Crisis in the Seattle-Tacoma, Washington, area*. ISCRAM, 2010.

Jordan Hochenbaum, Owen S Vallis, and Arun Kejariwal. Automatic Anomaly Detection in the Cloud via Statistical Learning. *arXiv preprint arXiv:1704.07706*, 2017.

Victoria Hodge and Jim Austin. A Survey of Outlier Detection Methodologies. *Artificial Intelligence Review*, 22(2):85–126, 2004.

Martin Rudi Holaker and Eirik Emanuelsen. Event Detection using Wikipedia. *124*, 2013.

Dawn E. Holmes, editor. *Innovations in Bayesian Networks: Theory and Applications.* Studies in Computational Intelligence. Springer-Verlag, Berlin Heidelberg, 2008. ISBN 978-3-540-85065-6.

Robin John Hyndman and George Athanasopoulos. Forecasting: Principles and Practice. 2018.

Anne Håkansson. Portal of Research Methods and Methodologies for Research Projects and Degree Projects. pages 67–73. CSREA Press U.S.A, 2013.

International Telecommunication Union. Statistics. https://www.itu.int/en/ITU-D/Statistics/Pages/stat/default.aspx. Accessed: 2019-04-03.

International Telecommunication Union. Digital Skills Toolkit. Technical report, International Telecommunication Union, Switzerland, Geneva, 2018.

Myungha Jang, Shiri Dori-Hacohen, and James Allan. Modeling Controversy Within Populations. In *Proceedings of the ACM SIGIR International Conference on Theory of Information Retrieval*, ICTIR '17, pages 141–149, New York, NY, USA, 2017. ACM. ISBN 978-1-4503-4490-6.

Sirkka L. Jarvenpaa, Thomas R. Shaw, and D. Sandy Staples. Toward Contextualized Theories of Trust: The Role of Trust in Global Virtual Teams. *Information Systems Research*, 15(3):250–267, September 2004. ISSN 1047-7047.

Bin Jiang. Head/tail Breaks: A New Classification Scheme for Data with a Heavy-tailed Distribution. *The Professional Geographer*, 65(3):482–494, August 2013. ISSN 0033-0124, 1467-9272. arXiv: 1209.2801.

Bin Jiang. Geospatial Analysis Requires a Different Way of Thinking: The Problem of Spatial Heterogeneity. *GeoJournal*, 80(1):1–13, February 2015a. ISSN 0343-2521, 1572-9893. arXiv: 1401.5889.

Bin Jiang. Head/tail Breaks for Visualization of City Structure and Dynamics. *Cities*, 43:69–77, March 2015b. ISSN 02642751. arXiv: 1501.03046.

Bin Jiang and Ding Ma. Defining Least Community as a Homogeneous Group in Complex Networks. *Physica A: Statistical Mechanics and its Applications*, 428:154–160, June 2015. ISSN 03784371. arXiv: 1502.00284.

Bin Jiang, Ding Ma, Junjun Yin, and Mats Sandberg. Spatial Distribution of City Tweets and Their Densities. *arXiv:1603.02231 [physics]*, March 2016. arXiv: 1603.02231.

Bin Jiang and Atsuyuki Okabe. Different Ways of Thinking about Street Networks and Spatial Analysis. *Geographical Analysis*, 46(4):341–344, 2014. ISSN 1538-4632.

Bin Jiang and Daniel Sui. A New Kind of Beauty Out of the Underlying Scaling of Geographic Space. *The Professional Geographer*, 66(4):676–686, October 2014. ISSN 0033-0124, 1467-9272. arXiv: 1303.7303.

Todd D. Jick. Mixing Qualitative and Quantitative Methods: Triangulation in Action. *Administrative Science Quarterly*, 24(4):602–611, 1979. ISSN 0001-8392.

R. Burke Johnson and Anthony J. Onwuegbuzie. Mixed Methods Research: A Research Paradigm Whose Time Has Come. *Educational Researcher*, 33(7):14–26, October 2004. ISSN 0013-189X.

R Burke Johnson, Anthony J Onwuegbuzie, and Lisa A Turner. Toward a Definition of Mixed Methods Research. *Journal of Mixed Methods Research*, 1(2):112–133, 2007.

Argyris Kalogeratos, Panagiotis Zagorisios, and Aristidis Likas. Improving Text Stream Clustering Using Term Burstiness and Co-burstiness. In *Proceedings of the 9th Hellenic Conference on Artificial Intelligence*, SETN '16, pages 16:1–16:9, New York, NY, USA, 2016. ACM. ISBN 978-1-4503-3734-2.

Gerald C. Kane. It's a Network, Not an Encyclopedia: A Social Network Perspective on Wikipedia Collaboration. *Academy of Management Proceedings*, 2009(1):1–6, August 2009. ISSN 0065-0668.

Gerald C. Kane and Sam Ransbotham. Research Note—Content and Collaboration: An Affiliation Network Approach to Information Quality in Online Peer Production Communities. *Information Systems Research*, 27(2):424–439, June 2016. ISSN 1047-7047, 1526-5536.

T. A. Kass-Hout and H. Alhinnawi. Social Media in Public Health. *British Medical Bulletin*, 108(1):5–24, December 2013. ISSN 0007-1420, 1471-8391.

Brian Keegan, Darren Gergle, and Noshir Contractor. 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, pages 105–113, New York, NY, USA, 2011. ACM. ISBN 978-1-4503-0909-7.

Brian Keegan, Darren Gergle, and Noshir Contractor. Staying in the Loop: Structure and Dynamics of Wikipedia's Breaking News Collaborations. In *Proceedings of the Eighth Annual International Symposium on Wikis and Open Collaboration*, WikiSym '12, pages 1:1–1:10, New York, NY, USA, 2012. ACM. ISBN 978-1-4503-1605-7. eventplace: Linz, Austria.

Brian C. Keegan and Jed R. Brubaker. 'Is' to 'Was': Coordination and Commemoration in Posthumous Activity on Wikipedia Biographies. In *Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work*, CSCW '15, pages 533–546, New York, NY, USA, 2015. ACM. ISBN 978-1-4503-2922-4.

P. P. Khaing and N. New. Adaptive Methods for Efficient Burst and Correlative Burst Detection. In *2017 IEEE/ACIS 16th International Conference on Computer and Information Science (ICIS)*, pages 453–458, May 2017.

A. U. R. Khan, M. B. Khan, and K. Mahmood. Cloud Service for Assessment of News' Popularity in Internet Based on Google and Wikipedia Indicators. In *2015 5th National Symposium on Information Technology: Towards New Smart World (NSITNSW)*, pages 1–8, February 2015.

Aniket Kittur, Bongwon Suh, Bryan A. Pendleton, and Ed H. Chi. He Says, She Says: Conflict and Coordination in Wikipedia. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '07, pages 453–462, New York, NY, USA, 2007. ACM. ISBN 978-1-59593-593-9.

C. Klemm, E. Das, and T. Hartmann. Swine Flu and Hype: A Systematic Review of Media Dramatization of the H1N1 Influenza Pandemic. *Journal of Risk Research*, 19 (1):1–20, 2016. ISSN 1366-9877.

George J. Klir. Principles of Uncertainty: What are They? Why Do we Need Them? *Fuzzy Sets and Systems*, 74(1):15–31, August 1995. ISSN 0165-0114.

George J. Klir. Generalized Information Theory: Aims, Results, and Open Problems. *Reliability Engineering & System Safety*, 85(1):21–38, July 2004. ISSN 0951-8320.

George J. Klir. *Uncertainty and Information: Foundations of Generalized Information Theory.* John Wiley & Sons, Inc., Hoboken, NJ, USA, November 2005. ISBN 978-0-471-75557-9 978-0-471-74867-0.

Leanne Knobloch and Kelly G. McAninch. Uncertainty Management. *Interpersonal Communication*, pages 297–319, January 2014.

Marina Kogan, Jennings Anderson, Leysia Palen, Kenneth M. Anderson, and Robert Soden. 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, pages 2783–2795, New York, NY, USA, 2016. ACM. ISBN 978-1-4503-3362-7.

Malcolm Koo. Complementary and Alternative Medicine on Wikipedia: Opportunities for Improvement, 2014.

Nikolaos Korfiatis and Ambjörn Naeve. *Evaluating Wiki Contributions Using Social Networks : A Case Study on Wikipedia.* Rinton Press, New Jersey, USA, 2005. ISBN 978-1-58949-050-5.

Kami Kosenko. Uncertainty Management Theory. In *Encyclopedia of Health Communication*, pages 1426–1427. SAGE Publications, Inc., Thousand Oaks, 2014.

Chakravanti Rajagopalachari Kothari. *Research Methodology: Methods and Techniques.* New Age International, 2004.

Shota Kotozaki. Identifying Local Burstiness in a Sequence of Batched Georeferenced Documents. *International Journal of Electronic Commerce Studies*, 6(2):269–288, December 2015. ISSN 20739729.

Natalie Kupferberg and Bridget McCrate Protus. Accuracy and Completeness of Drug Information in Wikipedia:An Assessment. *Journal of the Medical Library Association : JMLA*, 99(4):310–313, October 2011. ISSN 1536-5050.

Stacey Kuznetsov. Motivations of Contributors to Wikipedia. *SIGCAS Comput. Soc.*, 36(2), June 2006. ISSN 0095-2737.

Mirko Kämpf, Eric Tessenow, Dror Y. Kenett, and Jan W. Kantelhardt. The Detection of Emerging Trends Using Wikipedia Traffic Data and Context Networks. *PLOS ONE*, 10(12):e0141892, December 2015. ISSN 1932-6203.

David Laniado, Andreas Kaltenbrunner, Carlos Castillo, and Mayo Fuster Morell. Emotions and Dialogue in a Peer-production Community: The Case of Wikipedia. In *Proceedings of the Eighth Annual International Symposium on Wikis and Open Collaboration*, WikiSym '12, pages 9:1–9:10, New York, NY, USA, 2012. ACM. ISBN 978-1-4503-1605-7.

Theodoros Lappas, Benjamin Arai, Manolis Platakis, Dimitrios Kotsakos, and Dimitrios Gunopulos. On Burstiness-aware Search for Document Sequences. In *Proceedings of the 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '09, pages 477–486, New York, NY, USA, 2009. ACM. ISBN 978-1-60558-495-9.

Theodoros Lappas, Marcos R. Vieira, Dimitrios Gunopulos, and Vassilis J. Tsotras. STEM: A Spatio-temporal Miner for Bursty Activity. In *Proceedings of the 2013 ACM SIGMOD International Conference on Management of Data*, SIGMOD '13, pages 1021–1024, New York, NY, USA, 2013a. ACM. ISBN 978-1-4503-2037-5.

Theodoros Lappas, Marcos R. Vieira, Dimitrios Gunopulos, and Vassilis J. Tsotras. STEM: A Spatio-temporal Miner for Bursty Activity. In *Proceedings of the 2013 ACM SIGMOD International Conference on Management of Data*, SIGMOD '13, pages 1021–1024, New York, NY, USA, 2013b. ACM. ISBN 978-1-4503-2037-5.

M. R. Laurent and T. J. Vickers. Seeking Health Information Online: Does Wikipedia Matter? *Journal of the American Medical Informatics Association*, 16(4):471–479, July 2009. ISSN 1067-5027, 1527-974X.

Paul Lavrakas. z-Score. In *Encyclopedia of Survey Research Methods.* Sage Publications, Inc., 2455 Teller Road, Thousand Oaks California 91320 United States of America, 2008. ISBN 978-1-4129-1808-4 978-1-4129-6394-7.

Michael R Law, Barbara Mintzes, and Steven G Morgan. The Sources and Popularity of Online Drug Information: An Analysis of Top Search Engine Results and Web Page Views. *Annals of Pharmacotherapy*, 45(3):350–356, March 2011. ISSN 1060-0280.

Sungjun Lee, Sangjin Lee, Kwanho Kim, and Jonghun Park. Bursty Event Detection from Text Streams for Disaster Management. In *Proceedings of the 21st International Conference on World Wide Web*, WWW '12 Companion, pages 679–682, New York, NY, USA, 2012. ACM. ISBN 978-1-4503-1230-1.

Andreas Leithner, Werner Maurer-Ertl, Mathias Glehr, Joerg Friesenbichler, Katharina Leithner, and Reinhard Windhager. Wikipedia and Osteosarcoma: A Trustworthy Patients' Information? *Journal of the American Medical Informatics Association : JAMIA*, 17(4):373–374, 2010. ISSN 1067-5027.

Songnian Li, Suzana Dragicevic, Francesc Antón Castro, Monika Sester, Stephan Winter, Arzu Coltekin, Christopher Pettit, Bin Jiang, James Haworth, Alfred Stein, and Tao Cheng. Geospatial Big Data Handling Theory and Methods: A Review and Research Challenges. *ISPRS Journal of Photogrammetry and Remote Sensing*, 115:119–133, May 2016. ISSN 0924-2716.

Y. Li, J. Chen, and L. Feng. Dealing with Uncertainty: A Survey of Theories and Practices. *IEEE Transactions on Knowledge and Data Engineering*, 25(11):2463–2482, November 2013. ISSN 1041-4347.

JiYe Liang and YuHua Qian. Information Granules and Entropy Theory in Information Systems. *Science in China Series F: Information Sciences*, 51(10):1427–1444, October 2008. ISSN 1862-2836.

Kwan Hui Lim, Amitava Datta, and Michael Wise. A Preliminary Study on the Effects of Barnstars on Wikipedia Editing. In *Proceedings of the 9th International Symposium on Open Collaboration*, WikiSym '13, pages 23:1–23:2, New York, NY, USA, 2013. ACM. ISBN 978-1-4503-1852-5.

Yue Lin. *A Comparison Study on Natural and Head/tail Breaks Involving Digital Elevation Models*. 2013.

Ruoqian Liu, Ankit Agrawal, Wei-Keng Liao, and Alok Nidhi Choudhary. Enhancing Financial Decision-Making Using Social Behavior Modeling. In *Proceedings of the 8th Workshop on Social Network Mining and Analysis, SNAKDD 2014*, page a13. Association for Computing Machinery, Inc, January 2014.

Olimpia Lombardi, Federico Holik, and Leonardo Vanni. What is Shannon Information? *Synthese*, 193(7):1983–2012, July 2016. ISSN 1573-0964.

Xinjiang Lu, Zhiwen Yu, Bin Guo, Jiafan Zhang, Alvin Chin, Jilei Tian, and Yang Cao. Trending Words Based Event Detection in Sina Weibo. In *Proceedings of the 2014*

International Conference on Big Data Science and Computing, BigDataScience '14, pages 4:1–4:6, New York, NY, USA, 2014. ACM. ISBN 978-1-4503-2891-3.

Thomas Ludwig, Christoph Kotthaus, Christian Reuter, Sören van Dongen, and Volkmar Pipek. Situated Crowdsourcing During Disasters: Managing the Tasks of Spontaneous Volunteers through Public Displays. *International Journal of Human-Computer Studies*, 102:103–121, June 2017. ISSN 10715819.

Magnus Lundgren, Theresa Squatrito, and Jonas Tallberg. Stability and Change in International Policy-Making: A Punctuated Equilibrium Approach. *The Review of International Organizations*, 13(4):547–572, December 2018. ISSN 1559-744X.

Kalle Lyytinen and Mike Newman. Explaining Information Systems Change: a Punctuated Socio-Technical Change Model. *European Journal of Information Systems*, 17 (6):589–613, December 2008. ISSN 0960-085X.

Ding Ma, Mats Sandberg, and Bin Jiang. Characterizing the Heterogeneity of the OpenStreetMap Data and Community. *ISPRS International Journal of Geo-Information*, 4(2):535–550, June 2015.

Ding Ma, Mats Sandberg, and Bin Jiang. A Socio-Geographic Perspective on Human Activities in Social Media. *Geographical Analysis*, 49(3):328–342, 2017.

A.D. Madden. A Definition of Information. *Aslib Proceedings*, 52(9):343–349, November 2000. ISSN 0001-253X.

P. D. Magnus. On Trusting Wikipedia. *Episteme*, 6(01):74–90, February 2009. ISSN 1742-3600, 1750-0117.

Moira Maguire and Brid Delahunt. Doing a Thematic Analysis: A Practical, Step-by-step Guide for Learning and Teaching Scholars. *All Ireland Journal of Higher Education*, 9(3), October 2017. ISSN 2009-3160.

Naim Mahroum, Nicola Luigi Bragazzi, Kassem Sharif, Vincenza Gianfredi, Daniele Nucci, Roberto Rosselli, Francesco Brigo, Mohammad Adawi, Howard Amital, and Abdulla Watad. Leveraging Google Trends, Twitter, and Wikipedia to Investigate the Impact of a Celebrity's Death From Rheumatoid Arthritis. *Jcr: Journal of Clinical Rheumatology*, 24(4):188–192, June 2018. ISSN 1076-1608.

Silviu Maniu, Bogdan Cautis, and Talel Abdessalem. Building a Signed Network from Interactions in Wikipedia. In *Databases and Social Networks*, DBSocial '11, pages 19–24, New York, NY, USA, 2011. ACM. ISBN 978-1-4503-0650-8.

Michelle M. Maresh-Fuehrer and Richard Smith. Social Media Mapping Innovations for Crisis Prevention, Response, and Evaluation. *Computers in Human Behavior*, 54: 620–629, January 2016. ISSN 07475632.

H. Margetts, P. John, S. Hale, and T. Yasseri. *Political Turbulence: How Social Media Shape Collective Action.* Princeton Univ Press, Princeton, 2016. ISBN 978-1-4008-7355-5 978-0-691-15922-5.

P. McCauley-Bell and R. Freeman. Uncertainty Management in Information Warfare. In *Computational Cybernetics and Simulation 1997 IEEE International Conference on Systems, Man, and Cybernetics*, volume 2, pages 1942–1947 vol.2, October 1997.

Deborah L. McGuinness, Honglei Zeng, Paulo S. Pinheiro, Li Ding, Dhyanesh Narayanan, and Mayukh Bhaowal. Investigations into Trust for Collaborative Information Repositories: A Wikipedia Case Study. In *MTW*, 2006.

David J. McIver and John S. Brownstein. Wikipedia Usage Estimates Prevalence of Influenza-Like Illness in the United States in Near Real-Time. *PLOS Computational Biology*, 10(4):e1003581, April 2014. ISSN 1553-7358.

Márton Mestyán, Taha Yasseri, and János Kertész. Early Prediction of Movie Box Office Success Based on Wikipedia Activity Big Data. *PLOS ONE*, 8(8):e71226, August 2013. ISSN 1932-6203.

Sylvia L. Mikucki-Enyart, Sarah R. Petitte, and Sarah E. Wilder. Relational Uncertainty Management in Adult Children of Divorce. *Journal of Applied Communication Research*, 46(4):469–490, July 2018. ISSN 0090-9882.

Laura E. Miller. Sources of Uncertainty in Cancer Survivorship. *Journal of Cancer Survivorship*, 6(4):431–440, December 2012. ISSN 1932-2267.

Merle H. Mishel. Uncertainty in Chronic Illness. *Annual Review of Nursing Research*, 17(1):269–294, January 1999. ISSN 0739-6686, 1944-4028.

Helen Susannah Moat, Chester Curme, Adam Avakian, Dror Y. Kenett, H. Eugene Stanley, and Tobias Preis. Quantifying *Wikipedia* Usage Patterns Before Stock Market Moves. *Scientific Reports*, 3:1801, May 2013. ISSN 2045-2322.

Helen Susannah Moat, Chester Curme, H. Eugene Stanley, and Tobias Preis. Anticipating Stock Market Movements with Google and Wikipedia. In Davron Matrasulov and H. Eugene Stanley, editors, *Nonlinear Phenomena in Complex Systems: From Nano to Macro Scale*, NATO Science for Peace and Security Series C: Environmental Security, pages 47–59. Springer Netherlands, 2014. ISBN 978-94-017-8704-8.

Chris R. Morse. Theory of Motivated Information Management. In *Encyclopedia of Health Communication*, pages 1391–1392. SAGE Publications, Inc., Thousand Oaks, 2014.

Donald P. Moynihan. Learning under Uncertainty: Networks in Crisis Management. *Public Administration Review*, 68(2):350–365, 2008. ISSN 1540-6210.

Claudia Müller-Birn, Leonhard Dobusch, and James D. Herbsleb. Work-to-rule: The Emergence of Algorithmic Governance in Wikipedia. In *Proceedings of the 6th International Conference on Communities and Technologies*, C&T '13, pages 80–89, New York, NY, USA, 2013. ACM. ISBN 978-1-4503-2104-4.

Gordon Müller-Seitz. Practising Uncertainty in the Face of Large-Scale Disease Outbreaks. *Journal of Management Inquiry*, 23(3):276–293, July 2014. ISSN 1056-4926.

Michael Newman and Daniel Robey. A Social Process Model of User-Analyst Relationships. *MIS Quarterly*, 16(2):249–266, 1992. ISSN 0276-7783.

Sabine Niederer and José van Dijck. Wisdom of the Crowd or Technicity of Content? Wikipedia as a Sociotechnical System. *New Media & Society*, 12(8):1368–1387, December 2010. ISSN 1461-4448.

Khairul Baharein Moh Noor. Case Study: A Strategic Research Methodology. *American Journal of Applied Sciences*, 5(11):1602–1604, November 2008. ISSN 15469239.

N. A. Noori, J. J. Wolbers, F. K. Boersma, and X. V. Cardona. A Dynamic Perspective of Emerging Coordination Clusters in Crisis Response Networks. *ISCRAM 2016 Conference Proceedings – 13th International Conference on Information Systems for Crisis Response and Management*, 2016.

Lorelli S. Nowell, Jill M. Norris, Deborah E. White, and Nancy J. Moules. Thematic Analysis: Striving to Meet the Trustworthiness Criteria. *International Journal of Qualitative Methods*, 16(1):1609406917733847, December 2017. ISSN 1609-4069.

Chitu Okoli, Mohamad Mehdi, Mostafa Mesgari, Finn Årup Nielsen, and Arto Lanamäki. Wikipedia in the Eyes of its Beholders: A Systematic Review of Scholarly Research on Wikipedia Readers and Readership. *American Society for Information Science and Technology. Journal*, 65(12):2381–2403, 2014. ISSN 2330-1635.

Zina O'Leary. *The Essential Guide to Doing Research*. SAGE, June 2004. ISBN 978-0-7619-4199-6. Google-Books-ID: ItKeqNfgNW0C.

Florin Oprescu, Shelly Campo, John Lowe, Julie Andsager, and Jose A Morcuende. Online Information Exchanges for Parents of Children With a Rare Health Condition: Key Findings From an Online Support Community. *Journal of Medical Internet Research*, 15(1), January 2013. ISSN 1439-4456.

Leysia Palen, Kate Starbird, Sarah Vieweg, and Amanda Hughes. 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, June 2010. ISSN 1550-8366.

Leysia Palen, Sarah Vieweg, Jeannette Sutton, Sophia B Liu, and Amanda Hughes. Crisis Informatics: Studying Crisis in a Networked World. In *Third International Conference on e-Social Science*, volume 7. Citeseer, 2007.

Chul Hyun Park and Erik Johnston. Crowdsourced, Voluntary Collective Action in Disasters. In *Proceedings of the 16th Annual International Conference on Digital Government Research*, dg.o '15, pages 329–330, New York, NY, USA, 2015. ACM. ISBN 978-1-4503-3600-0.

Jennifer Phillips, Connie Lam, and Lisa Palmisano. Analysis of the Accuracy and Readability of Herbal Supplement Information on Wikipedia. *Journal of the American Pharmacists Association: JAPhA*, 54(4):406–414, August 2014. ISSN 1544-3450.

Jean Piaget. PiagetÄôs Theory. In B§rbel Inhelder, Harold H. Chipman, and Charles Zwingmann, editors, *Piaget and His School: A Reader in Developmental Psychology*, Springer Study Edition, pages 11–23. Springer Berlin Heidelberg, Berlin, Heidelberg, 1976. ISBN 978-3-642-46323-5.

Alicja Piotrkowicz, Vania Dimitrova, and Katja Markert. Automatic Extraction of News Values from Headline Text. *Proceedings of the Student Research Workshop at the 15th Conference of the European Chapter of the Association for Computational Linguistics*, pages 64–74, 2017.

Reid Priedhorsky, Dave Osthus, Ashlynn R. Daughton, Kelly R. Moran, Nicholas Generous, Geoffrey Fairchild, Alina Deshpande, and Sara Y. Del Valle. 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, 2017.

Keith Punch. *Introduction to Social Research : Quantitative and Qualitative Approaches*. Los Angeles, California : SAGE, third edition edition, 2014. ISBN 978-1-4462-4092-2.

Yan Qu, P. F. Wu, and Xiaoqing Wang. 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*, pages 1–11, January 2009.

Hoda Sepehri Rad and Denilson Barbosa. Identifying Controversial Articles in Wikipedia: A Comparative Study. In *Proceedings of the Eighth Annual International Symposium on Wikis and Open Collaboration*, WikiSym '12, pages 7:1–7:10, New York, NY, USA, 2012. ACM. ISBN 978-1-4503-1605-7.

Stephen A. Rains. Health Information Seeking and the World Wide Web: An Uncertainty Management Perspective. *Journal of Health Communication*, 19(11):1296–1307, November 2014. ISSN 1081-0730.

Stephen A. Rains and Riva Tukachinsky. Information Seeking in Uncertainty Management Theory: Exposure to Information About Medical Uncertainty and Information-Processing Orientation as Predictors of Uncertainty Management Success. *Journal of Health Communication*, 20(11):1275–1286, 2015. ISSN 1087-0415.

N. J. Reavley, A. J. Mackinnon, A. J. Morgan, M. Alvarez-Jimenez, S. E. Hetrick, E. Killackey, B. Nelson, R. Purcell, M. B. H. Yap, and A. F. Jorm. Quality of Information Sources about Mental Disorders: A Comparison of Wikipedia with Centrally Controlled Web and Printed Sources. *Psychological Medicine*, 42(8):1753–1762, August 2012. ISSN 1469-8978, 0033-2917.

Sheldon M. Ross. *Introduction to Probability Models, Ninth Edition*. Academic Press, Inc., Orlando, FL, USA, 2006. ISBN 978-0-12-598062-3.

Rahul Roy, Priya Seetharaman, Amitava Dutta, and Myshkin Ingawale. Network Analysis of User Generated Content Quality in Wikipedia. *Online Information Review*, 37 (4):602–619, August 2013. ISSN 1468-4527.

M. Anjali Sastry. Problems and Paradoxes in a Model of Punctuated Organizational Change. *Administrative Science Quarterly*, 42(2):237–275, 1997. ISSN 0001-8392.

Philipp Schaer, Philipp Mayr, and Peter Mutschke. Implications of Inter-Rater Agreement on a Student Information Retrieval Evaluation. *arXiv:1010.1824 [cs]*, October 2010. arXiv: 1010.1824.

Frank Schiller. Urban Transitions: Scaling Complex Cities Down to Human Size. *Journal of Cleaner Production*, 112:4273–4282, January 2016. ISSN 0959-6526.

Erich Schubert, Michael Weiler, and Hans-Peter Kriegel. SigniTrend: Scalable Detection of Emerging Topics in Textual Streams by Hashed Significance Thresholds. In *Proceedings of the 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '14, pages 871–880, New York, NY, USA, 2014. ACM. ISBN 978-1-4503-2956-9.

Savino Sciascia and Massimo Radin. What Can Google and Wikipedia Can Tell Us About a Disease? Big Data Trends Analysis in Systemic Lupus Erythematosus. *International Journal of Medical Informatics*, 107:65–69, November 2017. ISSN 1386-5056.

Hoda Sepehri-Rad and Denilson Barbosa. Identifying Controversial Wikipedia Articles Using Editor Collaboration Networks. *ACM Trans. Intell. Syst. Technol.*, 6(1):5:1–5:24, March 2015. ISSN 2157-6904.

Nigel Shadbolt, Kieron OÄôHara, David De Roure, and Dame Wendy Hall. *The Theory and Practice of Social Machines*. Lecture Notes in Social Networks. Springer International Publishing, 2019. ISBN 978-3-030-10888-5.

Thomas Shafee, Gwinyai Masukume, Lisa Kipersztok, Diptanshu Das, Mikael Häggström, and James Heilman. Evolution of Wikipedia's Medical Content: Past, Present and Future. *J Epidemiol Community Health*, 71(11):1122–1129, November 2017. ISSN 0143-005X, 1470-2738.

C. E. Shannon. A Mathematical Theory of Communication. *The Bell System Technical Journal*, 27(3):379–423, July 1948. ISSN 0005-8580.

Katriona Shea, Michael J. Tildesley, Michael C. Runge, Christopher J. Fonnesbeck, and Matthew J. Ferrari. Adaptive Management and the Value of Information: Learning Via Intervention in Epidemiology. *PLoS Biology*, 12(10):e1001970, October 2014. ISSN 1545-7885.

Ronald E. Shiffler. Maximum Z Scores and Outliers. *The American Statistician*, 42(1): 79–80, February 1988. ISSN 0003-1305.

Philipp Singer, Florian Lemmerich, Robert West, Leila Zia, Ellery Wulczyn, Markus Strohmaier, and Jure Leskovec. Why We Read Wikipedia. In *Proceedings of the 26th International Conference on World Wide Web*, WWW '17, pages 1591–1600, Republic and Canton of Geneva, Switzerland, 2017. International World Wide Web Conferences Steering Committee. ISBN 978-1-4503-4913-0.

Karanjit Singh and Dr Shuchita Upadhyaya. *Outlier Detection: Applications And Techniques*.

Paul R. Smart and Nigel R. Shadbolt. Social Machines. In Mehdi Khosrow-Pour, editor, *Encyclopedia of Information Science and Technology*, pages 6855–6862. IGI Global, August 2014. ISBN 978-1-4666-5888-2.

Paul R. Smart, Elena Simperl, and Nigel Shadbolt. A Taxonomic Framework for Social Machines. In Daniele Miorandi, Vincenzo Maltese, Michael Rovatsos, Anton Nijholt, and James Stewart, editors, *Social Collective Intelligence: Combining the Powers of Humans and Machines to Build a Smarter Society*, pages 51–85. Springer, October 2014.

Ilya M. Sobol. *A Primer for the Monte Carlo Method*. CRC Press, April 2018. ISBN 978-1-351-46957-9. Google-Books-ID: rrBXDwAAQBAJ.

Kate Starbird. Digital Volunteerism During Disaster: Crowdsourcing Information Processing. pages 7–12, 2011.

Kate Starbird, Jim Maddock, Mania Orand, Peg Achterman, and Robert Mason. Rumors, false flags, and digital vigilantes: misinformation on Twitter after the 2013 Boston Narathon Bombing. In *iConference 2014 Proceedings*. iSchools, March 2014. ISBN 978-0-9884900-1-7.

Kate Starbird, Emma Spiro, Isabelle Edwards, Kaitlyn Zhou, Jim Maddock, and Sindhuja Narasimhan. 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, pages 360–371, New York, NY, USA, 2016. ACM. ISBN 978-1-4503-3362-7.

Thomas Steiner. Bots vs. Wikipedians, Anons vs. Logged-Ins (Redux): A Global Study of Edit Activity on Wikipedia and Wikidata. In *Proceedings of The International Symposium on Open Collaboration*, OpenSym '14, pages 25:1–25:7, New York, NY, USA, 2014. ACM. ISBN 978-1-4503-3016-9.

Margaret Stewart and B Gail Wilson. The Dynamic Role of Social Media During Hurricane Sandy: An Introduction of the STREMII Model to Weather the Storm of the Crisis Lifecycle. 54, 07 2015.

Christopher T. Street and Darren B. Meister. Small Business Growth and Internal Transparency: The Role of Information Systems. *MIS Quarterly*, 28(3):473–506, 2004. ISSN 0276-7783.

B. Suh, E. H. Chi, B. A. Pendleton, and A. Kittur. Us vs. Them: Understanding Social Dynamics in Wikipedia with Revert Graph Visualizations. In *2007 IEEE Symposium on Visual Analytics Science and Technology*, pages 163–170, October 2007.

Bongwon Suh, Ed H. Chi, Aniket Kittur, and Bryan A. Pendleton. Lifting the Veil: Improving Accountability and Social Transparency in Wikipedia with Wikidashboard. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '08, pages 1037–1040, New York, NY, USA, 2008. ACM. ISBN 978-1-60558-011-1.

Bongwon Suh, Gregorio Convertino, Ed H. Chi, and Peter Pirolli. The Singularity is Not Near: Slowing Growth of Wikipedia. In *Proceedings of the 5th International Symposium on Wikis and Open Collaboration*, WikiSym '09, pages 8:1–8:10, New York, NY, USA, 2009. ACM. ISBN 978-1-60558-730-1.

Róbert Sumi, András Rung, András Kornai, and János Kertész. Characterization and Prediction of Wikipedia Edit Wars. pages 1–3, Koblenz, Németország, June 2011.

Michael Sunnafrank. Predicted Outcome Value and Uncertainty Reduction Theories A Test of Competing Perspectives. *Human Communication Research*, 17(1):76–103, September 1990. ISSN 1468-2958.

Miroslav Svtek. Information Physics and Complex Information Systems. *Encyclopedia of Information Science and Technology, Third Edition*, pages 7450–7455, 2015.

Jason Swarts. The Collaborative Construction of "Fact" on Wikipedia. In *Proceedings of the 27th ACM International Conference on Design of Communication*, SIGDOC '09, pages 281–288, New York, NY, USA, 2009. ACM. ISBN 978-1-60558-559-8.

Laura P. Swiler, Thomas L. Paez, and All L. Mayes. Epistemic Uncertainty Quantification Tutorial. In *Proceedings of the IMAC XXVII Conference and Exposition on Structural Dynamics*, 2009.

Jian Tang, Zhixiang Chen, Ada Waichee Fu, and David W Cheung. Capabilities of Outlier Detection Schemes in Large Datasets, Framework and Methodologies. *Knowledge and Information Systems*, 11(1):45–84, 2007.

Yla Tausczik, Kate Faasse, James W. Pennebaker, and Keith J. Petrie. Public Anxiety and Information Seeking Following the H1n1 Outbreak: Blogs, Newspaper Articles, and Wikipedia Visits. *Health Communication*, 27(2):179–185, February 2012. ISSN 1041-0236, 1532-7027.

Norman J. Temple and Joy Fraser. How Accurate Are Wikipedia Articles in Health, Nutrition, and Medicine? *Canadian Journal of Information and Library Science*, 38 (1):37–52, June 2014. ISSN 1920-7239.

Nguyen Thanh and Tran Thanh. The Interconnection Between Interpretivist Paradigm and Qualitative Methods in Education. *American Journal of Educational Science*, 2015.

Marijn ten Thij, Yana Volkovich, David Laniado, and Andreas Kaltenbrunner. Modeling Page-View Dynamics on Wikipedia. *arXiv:1212.5943 [physics]*, December 2012. arXiv: 1212.5943.

Tim Berners-Lee. "The Two Magics of Web Science" 10 Years On, 2007.

Ramine Tinati, Susan Halford, Leslie Carr, and Catherine Pope. Mixing Methods and Theory to Explore Web Activity. In *Proceedings of the 4th Annual ACM Web Science Conference*, WebSci '12, pages 308–316, New York, NY, USA, 2012. ACM. ISBN 978-1-4503-1228-8. event-place: Evanston, Illinois.

Ramine Tinati and Markus Luczak-Roesch. Wikipedia: A Complex Social Machine by Ramine Tinati and Markus Luczak-Roesch with Martin Vesely As Coordinator. *SIGWEB Newsl.*, (Winter):6:1–6:10, February 2017. ISSN 1931-1745.

Ramine Tinati, Markus Luczak-Roesch, Nigel Shadbolt, and Wendy Hall. Using WikiProjects to Measure the Health of Wikipedia. In *Proceedings of the 24th International Conference on World Wide Web*, WWW '15 Companion, pages 369–370, New York, NY, USA, 2015. ACM. ISBN 978-1-4503-3473-0.

Ramine Tinati, Thanassis Tiropanis, and Lesie Carr. An Approach for Using Wikipedia to Measure the Flow of Trends Across Countries. In *Proceedings of the 22Nd International Conference on World Wide Web*, WWW '13 Companion, pages 1373–1378, New York, NY, USA, 2013. ACM. ISBN 978-1-4503-2038-2.

Milena Tsvetkova, Ruth García-Gavilanes, Luciano Floridi, and Taha Yasseri. Even Good Bots Fight: The Case of Wikipedia. *PLOS ONE*, 12(2):e0171774, February 2017. ISSN 1932-6203.

Turi Machine Learning Platform User Guide. Local Outlier Factor, 2019.

Ralph H Turner, Lewis M Killian, et al. *Collective behavior*. Prentice-Hall Englewood Cliffs, NJ, 1957.

Besiki Stvilia Michael B. Twidale. Assessing Information Quality of a Community-Based Encyclopedia. In *Proceedings of the International Conference on Information Quality*, pages 442–454, 2005.

Inc. Twitter. Anomaly Detection with R. Contribute to Twitter/Anomaly Detection Development by Creating an Account on GitHub, March 2019a. original-date: 2014-12-09T17:46:24Z.

Inc. Twitter. Breakout Detection via Robust E-Statistics. Contribute to Twitter/BreakoutDetection Development by Creating an Account on GitHub, February 2019b. original-date: 2014-10-24T15:22:35Z.

Marlon Twyman, Brian C. Keegan, and Aaron Shaw. Black Lives Matter in Wikipedia: Collaboration and Collective Memory around Online Social Movements. *Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing - CSCW '17*, pages 1400–1412, 2017. arXiv: 1611.01257.

Mojtaba Vaismoradi, Hannele Turunen, and Terese Bondas. Content Analysis and Thematic Analysis: Implications for Conducting a Qualitative Descriptive Study. *Nursing & Health Sciences*, 15(3):398–405, September 2013. ISSN 1442-2018.

Andrew H. Van De Ven and Marshall Scott Poole. Explaining Development and Change in Organizations. *Academy of Management Review*, 20(3):510–540, July 1995. ISSN 0363-7425, 1930-3807.

Frans H Van Eemeren and Bart Garssen. *Controversy and Confrontation: Relating Controversy Analysis with Argumentation Theory*, volume 6. John Benjamins Publishing, 2008.

Tommaso Venturini. Diving in Magma: How to Explore Controversies with Actor-Network Theory. *Public Understanding of Science*, 19(3):258–273, May 2010. ISSN 0963-6625.

Sarah Vieweg, Amanda L. Hughes, Kate Starbird, and Leysia Palen. 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, pages 1079–1088, New York, NY, USA, 2010. ACM. ISBN 978-1-60558-929-9.

Santosh Vijaykumar, Yan Jin, and Claudia Pagliari. Outbreak Communication Challenges when Misinformation Spreads on Social Media. *Revista Eletrônica de Comunica\ccão, Informa\ccão e Inova\ccão em Saúde (RECIIS)*, 2019. ISSN 1981-6278.

Santosh Vijaykumar, Rianne Wally Meurzec, Karthikayen Jayasundar, Claudia Pagliari, and Yohan Fernandopulle. What's Buzzing on Your Feed? Health Authorities' Use of Facebook to Combat Zika in Singapore. *Journal of the American Medical Informatics Association*, 24(6):1155–1159, November 2017. ISSN 1067-5027.

Fernanda B. Viégas, Martin Wattenberg, and Kushal Dave. Studying Cooperation and Conflict Between Authors with History Flow Visualizations. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '04, pages 575–582, New York, NY, USA, 2004. ACM. ISBN 978-1-58113-702-6.

Michail Vlachos, Christopher Meek, Zografoula Vagena, and Dimitrios Gunopulos. Identifying Similarities, Periodicities and Bursts for Online Search Queries. In *Proceedings of the 2004 ACM SIGMOD International Conference on Management of Data*, SIGMOD '04, pages 131–142, New York, NY, USA, 2004. ACM. ISBN 978-1-58113-859-7.

Ba-Quy Vuong, Ee-Peng Lim, Aixin Sun, Minh-Tam Le, Hady Wirawan Lauw, and Kuiyu Chang. On Ranking Controversies in Wikipedia: Models and Evaluation. In *Proceedings of the 2008 International Conference on Web Search and Data Mining*, WSDM '08, pages 171–182, New York, NY, USA, 2008. ACM. ISBN 978-1-59593-927-2.

Claudia Wagner, Olga Zagovora, Tatiana Sennikova, and Fariba Karimi. Collective Attention Towards Scientists and Research Topics. In *Proceedings of the 10th ACM Conference on Web Science*, WebSci '18, pages 11–15, New York, NY, USA, 2018. ACM. ISBN 978-1-4503-5563-6.

Nicholas Walliman. *Research Methods: The Basics*. 01 2011. ISBN 9780415489942.

Canhui Wang, Min Zhang, Liyun Ru, and Shaoping Ma. Automatic Online News Topic Ranking Using Media Focus and User Attention Based on Aging Theory. In *Proceedings of the 17th ACM Conference on Information and Knowledge Management*, CIKM '08, pages 1033–1042, New York, NY, USA, 2008. ACM. ISBN 978-1-59593-991-3.

Abdulla Watad, Nicola Luigi Bragazzi, Francesco Brigo, Kassem Sharif, Howard Amital, Dennis McGonagle, Yehuda Shoenfeld, and Mohammad Adawi. Readability of Wikipedia Pages on Autoimmune Disorders: Systematic Quantitative Assessment. *Journal of Medical Internet Research*, 19(7):e260, 2017. ISSN 1438-8871.

Kerri Wazny. Applications of Crowdsourcing in Health: An Overview. *Journal of Global Health*, 8(1), 2018. ISSN 2047-2978.

P. Wei and N. Wang. *Wikipedia and Stock Return: Wikipedia Usage Pattern Helps to Predict the Individual Stock Movement*. Association for Computing Machinery, 2016. ISBN 978-1-4503-4144-8.

Howard T. Welser, Dan Cosley, Gueorgi Kossinets, Austin Lin, Fedor Dokshin, Geri Gay, and Marc Smith. Finding Social Roles in Wikipedia. In *Proceedings of the 2011 iConference*, iConference '11, pages 122–129, New York, NY, USA, 2011. ACM. ISBN 978-1-4503-0121-3.

Andrew G. West, Jian Chang, Krishna K. Venkatasubramanian, and Insup Lee. Trust in Collaborative Web Applications. *Future Generation Computer Systems*, 28(8):1238–1251, October 2012. ISSN 0167-739X.

O. G. Woodberry, K. B. Korb, and A. E. Nicholson. Testing Punctuated Equilibrium Theory Using Evolutionary Activity Statistics. In Kevin Korb, Marcus Randall, and Tim Hendtlass, editors, *Artificial Life: Borrowing from Biology*, Lecture Notes in Computer Science, pages 86–95. Springer Berlin Heidelberg, 2009. ISBN 978-3-642-10427-5.

Samuel Workman, Bryan D. Jones, and Ashley E. Jochim. Information Processing and Policy Dynamics. *Policy Studies Journal*, 37(1):75–92, February 2009. ISSN 1541-0072.

World Health Organization. NCDs | Noncommunicable diseases and their risk factors. http://www.who.int/ncds/en/, 2019a. Accessed: 2019-03-15.

World Health Organization. Non Communicable Diseases. https://www.who.int/news-room/fact-sheets/detail/noncommunicable-diseases, 2019b. Accessed: 2019-03-15.

World Health Organization. WHO | Disease Outbreaks. http://www.who.int/emergencies/diseases/en/, 2019c. Accessed: 2019-03-15.

Zheng Yi Wu, Yekun He, Qiao Li, et al. Comparing Deep Learning with Statistical Control Methods for Anomaly Detection. In *WDSA/CCWI Joint Conference Proceedings*, volume 1, 2018.

Thomas Wöhner, Sebastian Köhler, and Ralf Peters. Automatic Reputation Assessment in Wikipedia. *ICIS 2011 Proceedings*, December 2011.

Taha Yasseri and Jonathan Bright. Can Electoral Popularity be Predicted Using Socially Generated Big Data? *arXiv:1312.2818 [physics]*, December 2013. arXiv: 1312.2818.

Taha Yasseri and Jonathan Bright. Wikipedia Traffic Data and Electoral Prediction: Towards Theoretically Informed Models. *EPJ Data Science*, 5(1):22, June 2016. ISSN 2193-1127.

Taha Yasseri, Anselm Spoerri, Mark Graham, and János Kertész. The Most Controversial Topics in Wikipedia: A Multilingual and Geographical Analysis. *arXiv:1305.5566 [physics]*, May 2013. arXiv: 1305.5566.

Taha Yasseri, Robert Sumi, András Rung, András Kornai, and János Kertész. Dynamics of Conflicts in Wikipedia. *PLOS ONE*, 7(6):e38869, June 2012. ISSN 1932-6203.

Robert Yin. *Case Study Research: Design and Methods*. Fifth edition. edition, 2014. ISBN 978-1-4522-4256-9.

Robert K Yin. *Case Study Research: Design and Methods.* Thousand Oaks, Calif: Sage Publications, 3rd ed edition, 2003. ISBN 076192552X. Includes bibliographical references (pages 167-174) and index.

Zaidah Zainal. Case Study as a Research Method. *Jurnal Kemanusiaan*, 5(1), 2007.

Ark Fangzhou Zhang, Danielle Livneh, Ceren Budak, Lionel P. Robert, Jr., and Daniel M. Romero. Crowd Development: The Interplay Between Crowd Evaluation and Collaborative Dynamics in Wikipedia. *Proc. ACM Hum.-Comput. Interact.*, 1 (CSCW):119:1–119:21, December 2017. ISSN 2573-0142.

Ark Fangzhou Zhang, Ruihan Wang, Eric Blohm, Ceren Budak, Lionel P. Robert Jr, and Daniel M. Romero. Participation of New Editors after Times of Shock on Wikipedia. *Proceedings of the International AAAI Conference on Web and Social Media*, 13(01): 560–571, July 2019. ISSN 2334-0770.

Yan Zhang and Barbara M Wildemuth. Qualitative Analysis of Content. *Applications of Social Research Methods to Questions in Information and Library Science*, 308:319, 2009.

Wayne Xin Zhao, Baihan Shu, Jing Jiang, Yang Song, Hongfei Yan, and Xiaoming Li. Identifying Event-related Bursts via Social Media Activities. In *Proceedings of the 2012 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning*, EMNLP-CoNLL '12, pages 1466–1477, Stroudsburg, PA, USA, 2012. Association for Computational Linguistics.

Liu Zhiyong, Wang Qingyin, and Li Qi. Universal Information of Uncertainty Systematic Theory. In *2007 IEEE International Conference on Grey Systems and Intelligent Services*, pages 497–503, November 2007.

Mohammad Zohrabi. Mixed Method Research : Instruments , Validity , Reliability and Reporting Findings. 2013.