Using Twitter data to provide qualitative insights into pandemics and epidemics

By:

Wasim Ahmed

A thesis submitted in partial fulfilment of the requirements for the degree of Doctor of Philosophy

The University of Sheffield
Faculty of Social Sciences
Information School

January 2018

Abstract

Background: One area of public health research specialises in examining public views and opinions surrounding infectious disease outbreaks. Although interviews and surveys are valid sources of this information, views and opinions are necessarily generated by the context, rather than spontaneous. As such, social media has increasingly been viewed as legitimate source of pragmatic, unfiltered public opinion.

Objectives: This research attempts to better understand how users converse about infectious disease outbreaks on the social media platform Twitter. The study was undertaken in order to address a gap in knowledge because previous empirical studies that have analysed infectious disease outbreaks on Twitter have focused on employing quantitative methods as the primary form of data analysis. After analysing individual cases on Ebola, Zika, and swine flu, the study performs an important comparison in the types of discussions taking place on Twitter and is the first empirical study to do so.

Methods: A number of pilot studies were initially designed and conducted in order to help inform the main study. The study then manually labels tweets on infectious disease outbreaks assisted by the qualitative analysis programme NVivo, and performs an analysis using the Health Belief Model, concepts around information theory, and a number of sociological principles. The data were purposively sampled according to when Google Trends Data showed a heightened interest in the respective outbreaks, and a case study approach was utilised.

Results: A substantial number of themes were uncovered which were not reported in previous literature, demonstrating the potential of qualitative methodologies for extracting greater insight into public health opinions from Twitter data. The study noted several limitations of Twitter data for use in qualitative research. However, results demonstrated the potential of Twitter to identify discussions around infectious diseases that might not emerge in an interview and/or which might not be included in a survey.

Acknowledgements

There are a number of people who have contributed and supported me through this research process, with particular thanks to:

- My friends and family who have been a source of endless support through the research process. Especially my nephews who helped take my mind off the thesis when it was needed.

- My supervisors Peter Bath, Laura Sbaffi, and Gianluca Demartini for continual support and guidance without whom this project would not have been possible.

- Colleagues and friends across the Information School and the University of Sheffield, and in particular tutors on the *Researching Social Media* module which I undertook in my first year.

- The Faculty of Social Sciences at the University of Sheffield for providing a scholarship, and to my father Iftikhar Ahmed for providing additional financial and moral support.

- Those who kindly volunteered their time for the inter-coder reliability process Dimitrinka Atanasova, and Marc Bonne and also Nisar Ahmed for the company during the long hours in the library.

- Those who trusted me to talk at events and write articles on social media research.

- To all my A-Level teachers who supported me when I was predicted grades of DEE (because the area I lived in and previous grades). I would go on the achieve results of AAB. In particular, I would like to thank Janet Pike my Psychology A-level teacher for much inspiration during the time and for keeping in touch over the years.

REFEREED PUBLICATIONS ARISING FROM THIS WORK

Ahmed, W., Bath, P.A., Sbaffi, L., & Demartini, G. (2018). Moral Panic through the Lens of Twitter: An Analysis of Infectious Disease Outbreaks. *ACM Digital Library.* 9th International Conference on Social Media & Society. Copenhagen, Denmark, July-18-20, 2018.

Ahmed, W., Bath, P.A., Sbaffi, L., & Demartini. G. (2018). Measuring the Effect of Public Health Campaigns on Twitter: The Case of World Autism Awareness Day. *Lecture Notes of Computer Science* 2018. iConference Proceedings. Sheffield, United Kingdom, March 2018.

Ahmed, W., Bath, P.A., Sbaffi, L., & Demartini. G. (2018) Using Twitter for Insights into the 2009 Swine Flu and 2014 Ebola outbreaks. *IDEALS Proceedings*. iConference Proceedings. Sheffield, United Kingdom, March 2018.

Ahmed, W., Bath, P.A., & Demartini G (2017) Using Twitter as a data source: An overview of ethical challenges. Advances in Research Ethics and Integrity (Eds). Emerald Books.

Ahmed, W., Demartini, G., and Bath, PA. (2017) Topics Discussed on Twitter at the Beginning of the 2014 Ebola Epidemic in United States. iConference proceedings 2017. Wuhan, China, March 2017.

Ahmed, W., & Bath, PA. (2015) The Ebola epidemic on Twitter: challenges for health informatics. The Seventeenth International Symposium for Health Information Management proceedings (pp 289-289). York, 24 June 2015 – 26 June 2015.

Ahmed, W., & Bath, PA. (2015) Comparison of Twitter APIs and tools for analysing Tweets related to the Ebola Virus Disease. iFutures prooceedings. Sheffield, 07 July 2015.

KEY NON-PEER REVIEWED PUBLICATIONS

Ahmed, W. (2017). Review of the book The ethics of memory in a digital age: interrogating the right to be forgotten, by Ghezzi, A., Pereira, Â., & Vesnic-Alujevic, L. (Eds.). Information, Communication & Society. http://dx.doi.org/10.1080/1369118X.2017.1362456.

Ahmed, W. (2017) Using Twitter as a data source: An overview of current social media research tools (Updated for 2017) *LSE Impact of Social Sciences blog* http://blogs.lse.ac.uk/impactofsocialsciences/2017/05/08/using-twitter-as-a-data-source-an-overview-of-social-media-research-tools-updated-for-2017 8 May 2017 (Accessed 24/06/2017).

Ahmed, W. (2015) Challenges of using Twitter as a data source: An overview of current resources *LSE Impact of Social Sciences blog* http://blogs.lse.ac.uk/impactofsocialsciences/2015/09/28/challenges-of-using-twitter-as-a-data-source-resources/ 28 September 2015 (Accessed 10/10/2015).

Ahmed, W. (2015) Using Twitter as a data source: An overview of current social media research tools LSE Impact of Social Sciences blog http://blogs.lse.ac.uk/impactofsocialsciences/2015/07/10/social-media-research-tools-overview/ 10 July 2015 (Accessed 11/07/2015).

SELECTION OF INVITED TALKS

Ahmed, W. (2017). Keynote Talk – Gaining Powerful Insights into Social Media Listening. Boston University College of Communication. Making Social Media Data Matter. 20th Oct, 2017.

Ahmed, W., Bath, PA. (2017). Ethical Challenges of Social Media Data: Insights from Academia and Industry. OAI10 – CERN – UNIGE Workshop on Innovations in Scholarly Communication, Geneva, Switzerland.

Ahmed, W. (2017). Communicating Science Through Social Media: Tools and Techniques. *Society of Spanish Researchers in the UK.* Sheffield, 13th May 2017.

Ahmed, W. (2017). An Introduction to Social Network Analysis (Methodology) with NodeXL. Leeds Becket University, Leeds, 9th Jan 2017. (In association with the British Sociological Association).

Ahmed, W. (2017). Ethical Challenges of Using Social Media Data in Research. Bite Size Guide to Research in the 21st Century. ScHARR, Sheffield, 24th Jan 2017.

Ahmed, W. (2016). Insights from Social Media. Creative Entrepreneur, Media City (ITV, BBC HQs), Salford. 23 November 2016.

Ahmed, W. (2016). Introduction to NodeXL. An Introduction to Tools for Social Media Research. London, 11th October 2016.

Ahmed, W. (2016) Social Media Analytics. Department for Work and Pensions (DWP), Social Media Research Seminars. London, 4th August 2016.

Ahmed, W. (2016) workshop on Twitter Analytics at the Contemporary Issues In Economy & Technology (CIET- workshop for Nestle Adriatic). Split, 16-18th June 2016.

SELECTION OF SOCIAL MEDIA WORKSHOPS AND CONFERENCE PRESENTATIONS

Ahmed, W. (2017). Theoretical and Practical Foundations of Social Media Research (Keynote Talk). Gaining Powerful Insights into Social Media Listening: A Summer School. Vodice, Croatia. 28th June 2017.

Ahmed, W. (2017). One Minute Madness Poster Session. OAI10 – CERN – UNIGE Workshop on Innovations in Scholar nmhnfesfrtly Communication, Geneva, Switzerland.

Ahmed, W. (2017). Social Media: A Practical Approach. *Researching Social Media: A Theoretical and Practical* Approach. Sheffield. 30th May 2017.

Ahmed, W. (2017). Topics Discussed on Twitter at the Beginning of the 2014 Ebola Epidemic in United States. *Information School Seminar Series*. Sheffield, 7th March 2017.

Ahmed, W. (2017). Social Networks and Infectious Disease Outbreaks: The Case of Swine Flu and Ebola. PubhD, Sheffield, 1st Feb 2017.

Ahmed, W. (2017). Interview with Rony Robinson on the baring all feature. BBC Radio Sheffield, Sheffield, 19th Jan 2017.

Ahmed, W., Demartini, G., & Bath, PA. (2016) Using Twitter as a data source: An overview of ethical challenges. Ethics and Social Media Conference. London, 21st March 2016.

Ahmed, W., & Bath, PA. (2016) The Role of Social Media for Humanitarian Assistance and Disaster Management. Sheffield Institute for International Development (SIID) 7th Annual Postgraduate Conference. Sheffield. 7th-8th April 2016.

Ahmed, W (2015). Introduction to software that can be used to capture and analyse Twitter data. Audio lecture on Master of Social Science, Western Sydney University,27th August 2015.

Ahmed, W**.** (2015). The Ebola epidemic on Twitter: challenges for health informatics. Information School, Alumni Reunion Event. 25 November 2015.

Table of Contents

List of Tables

List of Figures

List of Appendices

List of Abbreviations

A/h1n1 – The medical term given to the 1918 Spanish influenza virus

Flu – Short for 'influenza'

CDC – Refers to the 'Centers for Disease Control and Prevention' which is based in the United States

WHO – Refers to the 'World Health Organisation' based in Geneva

H1N1/09 – The medical name given to the 2009 swine flu virus outbreak

API – Refers to 'application programming interface' which is an access point for Twitter data

Chapter 1 Introduction

1.1 Introduction

With the increase in globalisation, arising through, for example, modern commercial air travel, the risk of a major infectious disease pandemic has increased (Smith, 2006), and one area of public health research is the study of how people communicate about these diseases. One emerging platform for public health research is the social media platform Twitter. This thesis presents a research project that investigated the types of information that were shared on Twitter during two major infectious disease outbreaks (swine flu from 2009, and Ebola from 2014), and which examined similarities and differences between the ways people responded. After the initial study was designed, an outbreak of the Zika virus occurred in 2016, and therefore a further empirical study was also conducted to explore the information shared with that of swine flu and Ebola. Infectious diseases pose a serious public health risk and affect people globally. Although interviews and surveys are valid sources of this information, views and opinions are necessarily generated by the context, rather than spontaneous. As such, social media has increasingly been viewed as legitimate source of pragmatic, unfiltered public opinion. Recent reports have indicated that data from social media may be used for public health planning and health promotion.

This chapter provides background information and context for this project, outlines the research questions, and describes the structure of the thesis. Section 1.2 provides information to the topic of the thesis, and provides insight into information on recent and historical infectious disease outbreaks such as the Black Death and Spanish influenza pandemic, and reports on research conducted on infectious disease outbreaks on the World Wide Web. Section 1.3 provides the rationale for the study. Section 1.4 provides the research aims and objectives for this study; section 1.5 considers the research questions; section 1.6 considers the outcomes of the thesis and provides details of the contributions of this study; section 1.7 outlines the personal motivations behind the study; and, finally, section 1.8 provides an outline of the structure of the thesis.

1.2 Background

Infectious disease outbreaks are a serious public health concern that can cause a high number of fatalities within a short time, and which account for 29 out of 96 causes of major human mortality, as reported by the World Health Organisation and the World Bank (Taylor, Latham, and Mark, 2001; Murray and Lopez, 1996). The Black Death, which occurred from 1346 to 1353, is estimated to have taken the lives of between one-fourth and three-fourths of the world's population across Europe and Asia and, in Europe only, at least 25 million people died from it, including half of the London population at the time (around 100,000 people) (Haensch, *et al.*, 2010). The Spanish influenza pandemic (the medical name given to this strain of virus is A/H1N1), which occurred from 1918 to 1920, infected 500 million people and claimed 50-100 million lives (equivalent to 3% to 5% of the world's population) (Taubenberger and Morens, 2006). Although the Spanish influenza virus took more lives than the First World War, the war over-shadowed the pandemic (Bostrom, 2008; Haensch, *et al.*, 2010). These historical cases highlight how future infectious virus outbreaks may have similar disastrous consequences for global population health. More recently, the swine flu pandemic (the medical name given to this strain of virus is H1N1/09) in 2009, for instance, infected between 43 and 89 million people and claimed between 8,870 and 18,300 lives (CDC, 2010). The H1N1/09 swine flu outbreak lead to comparisons with the Spanish influenza by the public and journalists (Nicolson, 2009; Honigsbaum, 2009), although the number of fatalities was much less.

Due to increasing globalisation in the 21st century through international trade and travel, the likelihood that a deadly infectious disease outbreak would spread from one country to another has increased (Smith, 2006). In 2003, for example, the severe acute respiratory syndrome (SARS) spread from China to across at least 37 countries worldwide in merely a matter of weeks (Wang, 2004; Smith, 2006). One of the negative outcomes of infectious disease outbreaks is the very high mortality rate, and another is the large economic impact that they can have on the global economy (Smith, 2006). For the SARS outbreak, for instance, it was found that the global macro-economic impact was estimated to be $30 to $100 billion, and which ranged from $3 to $10 million per infected case (Smith, 2006; Smith & Sommers, 2003). The costs were dispersed across a number of different sectors, and those that were hardest hit were the travel and tourism industries (Smith, 2006).

These events are likely to lead to public views and opinions towards emerging diseases and the events that take place during an outbreak, which may be expressed on the online world, as a space for people to share their thoughts. People started sharing health information online towards the latter part of the 20th century; for example, via personal websites, discussion forums and online communities. The last few years have seen a shift towards sharing information via social media, and have changed the ways people communicate health issues (Smailhodzic, Hooijsma, Boonstra, and Langley, 2016). Previous deadly outbreaks, such as that of the Spanish influenza virus, occurred without modern communication devices such as personal computers and mobile phones. The 2009 swine flu pandemic, the 2014 Ebola epidemic, and the 2016 Zika outbreak occurred in the age of social media platforms such as Twitter. This makes it possible to examine unfiltered public views and opinions shared during these outbreaks and, importantly, to study what aspects of health online communities choose to converse about.

Twitter allows its users to send 280-character text updates (originally limited to 140 characters from 2006 to September 2017 (Telegraph, 2017) in the English language, known as 'tweets', such as thoughts, feelings, activities, and opinions (Chew and Eysenbach, 2010). For public health research, Twitter provides the unique opportunity to examine unmoderated discussions and information shared during disease outbreak such as swine flu and Ebola. Sections 1.2.1, and 1.2.2 will provide an outline of the 2009 swine flu pandemic and the 2014 Ebola epidemic.

1.2.1 Swine Flu

The swine influenza (flu) Pandemic of 2009 outbreak began in April 2009 and originated from Mexico (Davis, 2015) and spread across the world because it was a new strain of flu and members of the public had no immunity to it (World Health Organisation, 2009a; NHS Choices, 2015). The United States Centres for Disease Control and Prevention (CDC) announced on the 21st of April 2009 that two patients from California had been infected by the swine flu virus and this would lead to preparations for a swine flu pandemic. Four weeks after the initial two reports in California, 41 countries reported diagnosed cases of the virus (Wang and Palese, 2009). There were an estimated 123,000 to 203,000 deaths due to swine flu from 1st April to 31st December 2009 (Simonsen *et al.*, 2013). The swine flu pandemic caused members of the public in the US to worry about events that were taking place, and there was also an increase in media reports at that time. Further background information on the swine flu outbreak is provided in Chapter 4: Swine Flu, specifically in sections 4.2 and 4.3.

1.2.2 Ebola virus

The 2014 outbreak of the Ebola virus was traced to Guinea in December 2013, and spread across West Africa. As of January 2016, there were 28,637 cases of Ebola across and 11,315 deaths (Gulland, 2016). The 2014 Ebola epidemic was the largest epidemic of Ebola ever recorded, and the number of cases outnumbered all of the previous cases combined (Frieden, Damon, Bell, Kenyon, and Nichol, 2014). In June 2014, Médecins Sans Frontières (MSF) noted that the outbreak was out of control, and in August 2014 the United Nations (UN) declared Ebola to be an international public health emergency (BBC News, 2016).

During infectious disease outbreaks there is the potential for people from non-medical backgrounds to share health information which could potentially be harmful. As such, during the 2014 Ebola epidemic, it was reported that medical misinformation on the cures of Ebola had taken a number of lives due to the rumour that salt water had the potential to cure Ebola (Social Media Ebola Hoax Causes Deaths, 2015; Ngade, Singer, Marcus, Lara, 2016). Further background information on Ebola is provided in Chapter 5, specifically in sections 5.2 and 5.3.

1.2.3 Zika Virus

Zika became a public concern in early 2016, when it first spread outside of Africa and Asia where the virus had formerly been restricted. This geographical expansion lead the World Health Organisation (WHO) to declare the outbreak a public health emergency of international concern (PHEIC) (Fauci and Morens, 2016). The Zika virus is linked to adverse consequences during pregnancy as well as birth defects such as microcephaly and brain defects (Brasil *et al.*, 2016). However, unlike previous infectious disease outbreaks, there appeared to be low public knowledge of the Zika outbreak in the United States, leading to the dissemination of health information such as mosquito-bite prevention notices and recommendations by public health authorities to avoid travelling to Zika affected areas (Brasil *et al.*, 2016). Further background information on Zika will be provided in Chapter 6, specifically in sections 6.2 and 6.3.

1.2.4 Social Media and Infectious Disease Outbreaks

The 2009 swine flu and 2014 Ebola outbreaks are two health-related events with the highest proportion of media coverage in the 21st century (McCandless, 2009). Google ranked ‘swine flu' as the fastest rising Web search query in Google News (Google Zeitgeist, 2009), and 'Ebola' was among the most searched terms in 2014 (Google Trends, 2017).

Social media websites provide a space where information related to individual experiences, stories and opinions can be shared by members of the public (Chew and Eysenbach 2010; Nelson, and Staggers, 2013). Social media platforms have the potential to improve the quality of healthcare experiences a person might have, because they provide access to information and social support (Nelson, and Staggers, 2013; Smailhodzic, Hooijsma, Boonstra, and Langley, 2016). As originally intended, social media platforms were used to communicate with friends and family and to talk about common interests and hobbies (Smailhodzic, Hooijsma, Boonstra, and Langley, 2016). However, people's use of social media has expanded to include health information exchange, and some of those who use social media may now do so to receive and share information related to health and similar issues (Fox, and Duggan, 2013; Thackeray, Crookston, and West 2013; Smailhodzic, Hooijsma, Boonstra, and Langley, 2016). As such, it is important to develop a better understanding of how users communicate on this platform and to examine the information that is shared by users in order to assess its reliability.

Twitter is a social media platform that grew in popularity for personal use among members of the public, and early research found that users would post about their daily activities, and would seek and share information on public issues (Java, Song, Finin, Tseng, 2007). Twitter has emerged as an important resource for public health research and is used in many studies to track trends and behaviours related to illnesses and conditions (Nelson, and Staggers, 2013; Lee, Decamp, Dredze, Chisolm, and Berger, 2014). Twitter can also be used to provide social support in order to motivate health behaviour change, and to promote healthy behaviours (Smailhodzic, Hooijsma, Boonstra, and Langley, 2016). Tweets contain a wealth of data, and the analysis and mining of such data can provide insights into public opinion and behaviour responses (Chew and Eysenbach, 2010).

Twitter provides public access to its data, which can hence be retrieved more easily compared to other social media platforms such as Facebook. Thus, Twitter data provides a unique opportunity to yield insights into how users of this social media platform communicate about health. Members of the public may exchange health-related information privately on Twitter (for example via direct messages) or they can do so publicly. When people communicate publicly on Twitter, it is possible to retrieve this data based on whether it matches certain keywords or hashtags. More background information on Twitter will be provided in Chapter 2, Literature Review, section 2.10.

Twitter is also a platform that allows for the sharing of news stories with links to the webpages (an overview of the features can be found in Chapter 2). Health organisations such as the World Health Organisation, Centres for Disease Control, NHS England, Public Health England, and BBC Health (to name a few) are all now providing health information via Twitter. Users on Twitter may share the information that these organisations disseminate by tweeting and retweeting it to their followers. This leads to rapid spread and cascading of information. News articles may motivate users to express their views towards them (for example, if a user feels afraid of something in response to an article they may indicate this in a tweet). Tweets contain a time-stamp indicating when they were first posted, can be created by anyone using an Internet connection, and are publicly available (Thelwall, 2012).

Twitter has been used to find and examine whether poll data can be correlated to tweets by using sentiment analysis (O'Connor and Balasubramanyan, 2010), and it has been used for earthquake detection by using semantic analysis and search queries such as 'earthquake' and 'shaking' (Sakaki, Okazaki, and Matsuo, 2010). Moreover, Twitter has been credited as being influential during the Arab Spring, which were a series of protests that took place across the Middle East and North African countries from 2010 (Howard *et al.*, 2011). Twitter has therefore been researched across a number of disciplines.

Previous research has examined Twitter in relation to health topics such as dementia (Robilard *et al.*, 2013), antibiotics (Scanfeld, Scanfeld and Larson, 2010), marijuana (Cavazos-Rehg *et al.*, 2015), sexual risk behaviours (Young, Rivers and Lewis, 2014), and vaccination sentiments (positive and negative views), (Salathé and Khandelwal, 2011). Although previous empirical research has examined Twitter content surrounding swine flu (Chew and Eysenbach, 2010; Signorini, Segre and Polgreen, 2011; Kostkova, Szomszor, and St. Louis, 2014), and Ebola (Jin *et al.*, 2014; Oluwafemi, Elia, and Rolf, 2014; Odlum and Yoon, 2015), no research has conducted an in-depth thematic analysis of how Twitter users respond during infectious disease outbreaks. (as highlighted in Literature Review Chapter Two section 2.12)

1.3 Rationale for Study

In this study, tweets surrounding outbreaks of the H1N1 virus (swine flu) in 2009, the Ebola virus in 2014, and the Zika virus from 2016 were retrieved and analysed, and then compared to each other to highlight similarities and differences. This section explains the rationale for carrying out the present study.

It is important to understand the types of content shared on Twitter because tweets are viewed widely across the Internet, and a subset of the global human population use the platform on a daily basis (About Twitter, n.d.; Twitter Q1 2017 Company Metrics, 2017). Twitter reports having 328 million monthly active users, with one billion unique monthly visits to sites embedded with tweets (About Twitter, n.d.; Twitter Q1 2017 Company Metrics, 2017). Moreover, It is important to study the types of content shared on Twitter during infectious disease outbreaks because pandemic diseases are ranked as one of the most global catastrophic risks facing human civilisation (Bostrom, 2008), and potential misunderstanding and misinformation could have major negative outcomes. Traditionally, printed media have been the main information source for people interested in general health information during a major pandemic; however, for the 2009 swine flu pandemic, it was reported that the Internet had become the most common source of information (Jones and Salathe, 2009).

The previous paragraph has made an argument for the importance of studying the content on Twitter to gain an understanding of how people communicate about disease outbreaks because many people use the platform. It is also interesting to reflect on the fact that Twitter has the potential to influence the health behaviours of those that use it, but also the health behaviours of the friends and family members of those using the platform, if Twitter users discuss this information with other people. Its potential reach is therefore much higher than the number of users it hosts.

Different disciplines may design and conduct research differently, for example, studies conducted using Twitter data from the field of computer science may lack a sociological basis and describe the results without relating them to theory. Other disciplines, for instance, studies in Philosophy may be purely theoretical. This is not to say either of these approaches are improper because studies from these fields are still likely to contribute to knowledge. A novel aspect of the sociological study described in this thesis is that it utilised in-depth qualitative methods and aimed to describe results in the context of Information Theory (i.e. Information Seeking) and Health Theory (i.e. the Health Belief Model). This allows for a deeper interpretation of the results, and provides a richer description of the themes. A further novel aspect of this study involves the comparison of how Twitter users responded to swine flu from 2009, and Ebola from 2014 which was further compared to the Zika outbreak from 2016.

Previous empirical work has noted that Twitter has the potential to provide insight into public views and opinions for infectious disease outbreaks. However, to the best of the researcher's knowledge no previous empirical research (Chew and Eysenbach, 2010; Signorini, Segre and

Polgreen, 2011; Signorini, Segre and Polgreen, 2011; Kostkova, Szomszor, and St. Louis, 2014; Oluwafemi, Elia, and Rolf, 2014; Jin *et al.*, 2014; Odlum and Yoon, 2015; Scheinfeld, Bernhardt, Wilcox, and Suran, 2015; Kalyanam, Velupillai, Doan, Conway, and Lanckriet, 2015; Towers *et al.*, 2015; Rodriguez-Morales, Castañeda-Hernández, and McGregor, 2014) examining infectious disease outbreaks on Twitter has critically considered the strengths of and limitations of using Twitter data to provide such insights, and whether Twitter data is an effective substitute for traditional qualitative methods such as surveys and/or interviews. This study is distinct from previous empirical work examining Ebola, Swine Flu and Zika because it applies an in-depth qualitative method (thematic analysis) to extract individual insights (Chapters 4, 5 and 6) on the outbreaks. The study then compares the reactions of Twitter users across the different outbreaks (Chapter 7) in order to identify commonalities and differences, and is the first evidence based research to do so.

As is explained in section 7.3, the findings of this study may help to better inform media, health organisations, and policy makers on what information the public requires, and what types of health messages should be shared on Twitter. Previous research has suggested that the government should do more to meet the information needs of citizens during epidemics, for example, by providing more information on what vaccinations are available and their possible side effects (Chew and Eysenbach, 2010; Hilton and Smith, 2010; Seale *et al.*, 2010; Teasdale and Yardley 2011). Therefore, Governments could disseminate accurate factual information about diseases this way; however, there is always the potential for other information (that might be inaccurate or based on personal opinion) to be disseminated on the platforms. Twitter facilitates the study of human responses to events in a setting not bounded by a research question, and captures the immediate thoughts of the public. It allows users to share open views and thoughts that may not necessarily emerge outside the platform.

1.4 Research Aim and Objectives

The overall aim of this thesis is to develop a better understanding on how people communicated on Twitter during 2009 swine flu pandemic and the 2014 Ebola epidemic, and to explore the commonalities and differences between Twitter responses to outbreaks. More specifically, the objectives are:

- **Objective 1:** To develop an understanding of the current state of knowledge in this area, and to uncover potential gaps in literature.

- **Objective 2:** To gain an in-depth qualitative overview into how users communicated on Twitter during the swine flu pandemic and the Ebola epidemic.

- **Objective 3:** To examine potential similarities and differences in user responses related to the swine flu outbreak and the Ebola outbreak.

- **Objective 4:** To examine the similarities and differences in user responses related to the Zika outbreak and the swine flu and Ebola outbreaks.

1.5 Research Questions

The research questions this project sought to address are:

Q.1. What does the analysis of social media data (Twitter) tell us about how people communicate about disease outbreaks?

Q2. What similarities and differences emerge when comparing how people respond to infectious disease outbreaks on Twitter?

Q3. What are the advantages and limitations of using Twitter to provide in-depth insights into how citizens communicate about infectious disease outbreaks?

Q.4 What characteristics have enabled Twitter to thrive in the field of health-related public research in comparison to other social media platforms?

1.6 Outcomes of PhD

The theoretical contribution of this thesis is discussed in the literature review (Chapter 2, Literature Review, Sections 2.11 to 2.14), where studies that have used Twitter as a primary data source to examine health-related research were brought together and synthesised and in the Conclusions (Chapter 8). The intended methodological contributions of this PhD can be found in the Methodology Chapter 3 (section 3.12), where a strategy for retrieving and analysing a large amount of Twitter data was provided. A further methodological contribution relates to the reflection of whether Twitter is a worthwhile platform for gaining in-depth qualitative analysis (Discussion Chapter 7, section 7.9).

The intended substantial contributions to knowledge arising from this PhD relate to the in-depth thematic analysis of the case studies that have been selected (Chapter 5 Swine Flu section 5.7, Chapter 6 Ebola section 6.6, and Chapter 6 Zika, section 5.5), which provides new knowledge on the types of content shared on Twitter during the 2009 swine flu, the 2014 Ebola, and 2016 Zika outbreaks. These contributions shed light on how people communicate about infectious disease outbreaks on Twitter.

A further theoretical contribution relates to the applicability of Information Theory, i.e. of concepts such as information seeking and health theory (Health Belief Model), to the interpretation of results (Discussion Chapter 7, section 7.5 and section 7.6). The study also applies sociological concepts such as the moral panic and the outbreak narrative to the results of this study (Discussion Chapter 7, section 7.7 and section 7.8).

The study also contributes to knowledge by assessing the potential of Twitter data for gaining in-depth insights into infectious disease outbreaks by reflecting on this process. These contributions to new knowledge are discussed further in Conclusions Chapter 7 (section 7.6).

The results of this project will be of potential interest to organisations such as the World Health Organization (WHO), the United Nations (UN), the National Health Service (NHS), and Department for Environment, Food & Rural Affairs (DEFRA). A number of working relationships have been formed with some of these organisations. There has been interest surrounding the general aspects of this research project from a number of organisations. An invited-talk was delivered to the Department for Work and Pensions (DWP) government office in London in 2016. A talk was also delivered in January 2017 to delegates at a workshop organised by the British Sociological Association (BSA), and at an event organised by the Sociological Research Association (SRA) in collaboration with NatCen Social Research. A further invited-talk was delivered at a CERN workshop (the European Organization for Nuclear Research) in Geneva in 2017. An opening keynote talk was also delivered in October 2017 to Boston University College of Communication. A number of guest lectures were delivered to Masters Students at the University of Sheffield, and also to students at Western Sydney Australia. A full overview of talks and lectures delivered can be found on page 13 Selection of Invited Talks, and page 14 with the section heading Selection of Social Media Workshops and Conference Presentations.

1.7 Personal motivation for this research

The researcher completed an MSc degree at the Information School in 2013, with a dissertation on the evaluation of the information quality of Norovirus websites. In various discussions with the supervisors, the researcher then developed a research proposal which would study the use of social media for sharing health-related information because this was a new and exciting area for health research.

Twitter has an open nature, which means that discussions around key events may be easily studied. Twitter could provide a way to avoid limitations such as participants not correctly recalling how they felt at the time of an outbreak (i.e. recall bias) or the challenges posed by interview bias (McKee, 2013). When the researcher discovered the number of monthly active users on Twitter, it became apparent that studying the platform had potential.

The researcher, originally from a humanities background, needed to develop technical skills in order to understand the practicalities and technical aspects of social media research. These skills were developed by reviewing and testing social media research tools. This led to the writing and publishing of a number of blog posts for the London School of Economics and Political Sciences (LSE) Impact Blog. The most popular of these blog posts (Ahmed, 2015) reviewed social media research tools for social scientists. This post received over 23,000 page views, was shared over a thousand times on Facebook and Twitter respectively, and has since been cited in conference papers, journal articles, and PhD theses (Russell and Johnston, 2015; Mear, 2016; Taylor-Smith, 2016; Fonseca and Alves, 2017; Arabghalizi and Rahdari, 2017). Deveoping these skills also developed the reputation of the researcher for undertaking research on social media, and led to invitations to write blogs, invitations to presentations, and to lead workshops on social media research.

1.8 Thesis Structure

This section describes the structure of the thesis, and the content of each chapter.

Chapter Two - Literature Review. This chapter reviews the key literature related to this topic and covers four important areas: an overview of information and social cognition models (section 2.4 and section 2.5. It then outlines a number of sociological concepts relevant to this study (section 2.6). The chapter then provides a comprehensive understanding of the state of current knowledge in this new and rapidly developing area (section 2.7 to section 2.14) and identifies gaps in research (section 2.14) and how this research intends to fill these gaps (section 2.15).

Chapter Three - Methodology. This chapter outlines the research philosophy (section 3.1 and section 3..2), the research methods (section 3.4 and section 3.5), provides an overview of the data sources used in this study (section 3.6). The chapter then provides an overview of data analysis techniques (section 3.7) and techniques related to the quality of research (section 3.8). It then outlines Application Programming Interfaces (APIs) (section 3.9) and provides results of a study which compares different Twitter APIs (section 3.10). It then outlines a pilot study conducted on World Autism Awareness day (section 3.11), and a pilot study on Ebola tweets (section 3.12). The chapter then outlines ethical issues encountered in the research (section 3.13), provides an overview of the data gathering and filtering strategies (section 3.14), the data analysis techniques (section 3.15). The chapter then provides insight into the validity and reliability measures applied in the study (section 3.16), how the cases will be compared (section 3.17) and finally provides a summary of the chapter (section 3,18).

Chapter Four - Swine Flu Case Study. This chapter provides further background information on the swine flu pandemic from 2009 (section 4.2), including a timeline of key events during the outbreak (section 4.3). The chapter then summarises the data selection procedures (section 4.4), provides a summary of the data analyses (section 4.5), the platform features (section 4.6), the results from the qualitative analyses (section 4.8), and from the quantitative analyses (section 4.8). The chapter then provides insight into the frequency distribution of themes (section 4.9), popular content (section 4.10), and results from reliability measures (section 4.12 to 4.13). The chapter then provides insight into influential Twitter users (section 4.14), provides evidence for themes occurring across the outbreak (section 4.15). It then discusses the results of the study (section 4.15), provides an overview of how the Health Belief Model was used (section 4.16), describes the limitations of the results (section 4.17), and provides a summary of findings (section 4.18).

Chapter Five - Ebola Case Study. This chapter provides background information on the Ebola outbreak (section 5.2), provides a summary of key events (section 5.3), an overview of the data collection procedures (section 5.4), and summarises the platform features (section 4.5). The chapter then provides results from the qualitative analysis (section 5.6), an overview of the frequency distribution of themes (section 5.7), TAG clouds of sub-themes (section 5.8), and the results of the reliability measures (section 5.9). The section then provides a summary of the dataset (section 5.10), such as popular content. The chapter then identifies influential Twitter users (section 5.11), and provides evidence regarding the themes occurred across the outbreak (section 5.12). The chapter then provides a discussion of the results (section 5.13), discusses the limitations of the study (section 5.14), and provides a summary of the results of the study (section 5.15).

Chapter Six – Zika Case Study. This chapter first provides background information on the Zika outbreak (section 6.2); it then details key events taking place at the time (section 6.3), and summarises the data that was collected and analysed (section 6.4). The chapter then provides the results of the analysis (section 6.5), the frequency distribution of themes (section 6.6), and an overview of the reliability measures (section 6.7). It then provides evidence for a number of themes which occurred throughout the outbreak period (section 6.8), discusses the results (section 6.9), provides an overview of limitations (section 6.10), and summarises the results of the study (section 6.11).

Chapter Seven – Discussion. This chapter provides further background to the study (section 7.2), compares results from Chapters Four and Five (section 7.3), provides guidance to those in the position of disseminating information during infectious disease outbreaks (section 7.4), discusses the utility of the Health Belief Model (section 7.5), and concepts related to Information Theory (section 7.6). The chapter then draws on literature from sociology such the concept of the moral panic (section 7.7), and the outbreak narrative (section 7.8), and reflects on the study overall by discussing the utility of Twitter for qualitative research in comparison to traditional methods such as surveys and interviews (section 7.9).

Chapter Eight – Conclusion. This chapter summarises the findings by relating back to the research questions (section 8.2), it discusses the implications of the findings (section 8.3), the limitations (section 8.4) and the strengthens of the research (section 8.5), the contribution to current knowledge (section 7.6), and then provides a number of suggestions for future research (section 8.7). The thesis then provides an overall summary of the present study (section 8.8).

1.9 Summary

This chapter has introduced the thesis as a whole, it has set out the context and provided the rationale for the study, research questions and objectives. Infectious disease outbreaks have the potential for high mortality, and social media platforms such as Twitter are important new communication tools that should be studied to develop a better understanding of how people are communicating on these platforms. Hence, this thesis is looking at the information shared and communicated among Twitter users at the time of three major infectious disease outbreaks that occurred in 2009 (swine flu), 2014 (Ebola), and 2016 (Zika). The chapter also provided an overview of the structure of the thesis. The next chapter will review the key scientific literature related to this topic in order to summarise current knowledge and identify gaps in knowledge. It will then outline concepts related to Information Theory, as well as a number of Social Cognition Models that will be applied to the results of this study.

Chapter 2 Literature Review

2.1 Introduction

Chapter 1 introduced the general background and context for this study, provided a rationale for the importance of the topic, provided an overview of the research aims and objectives, research questions, motivations for the research, expected outcomes, and finally provided an outline of the thesis structure. This chapter provides a review of the existing literature on Twitter in relation to health, information theory, and a number of social cognition models. Before undertaking a piece of original research, it is important to review existing literature critically and thoroughly. This will allow for the identification of some of the key ideas and methodologies within the field where a contribution to knowledge can be made (Hart, 1998). Hart (1998) noted that the ideas and works of others are likely to provide a researcher with a framework that their own work can build upon. These could be regarding the methodology of the research, the techniques for data collection, and the key concepts within the field. After examining current subject knowledge on an area it will be possible to identify the gaps in literature and anomalies that may be present in previous research (Hart, 1998). A brief description of the sections in the chapter is provided in section 2.3.

2.2 Identifying Literature

2.2.1 Sources used

The University of Sheffield's Primo Central search engine was used to locate literature, which incorporates over 17 journal databases, including Scopus, MEDLINE, Springer Link, arXiv, and Web of Science. Google Scholar's search engine was also employed, as well as citation analysis of journal articles (i.e., following up references in bibliographies). It was found that a large volume of literature on Twitter had been published from 2012-2017. Search terms were also entered into Google.

2.2.2 Search Terms and Strategy

For the more general aspects of the literature review, search queries relating to a topic of interest, (i.e. information needs, information seeking, etc.) were entered into the Primo

Central search engine as well as Google Scholar. These included "Information Theory", "Information Seeking", "Information Seeking" AND "Internet", "Health Information" AND "Internet", "Infectious Diseases" AND "Internet". To identify literature on infectious diseases, search terms included "Infectious Diseases", "Pandemic", "Epidemic", "Swine Flu", "Ebola", and "H1N1".

Search terms used in locating literature included "Twitter" alongside a health condition; for example, "swine flu" AND "Twitter", or "Ebola" AND "Twitter", or "Zika" AND "Twitter" and as general as "Twitter" AND "health". For crisis and risk communication literature, the search terms were as simple as "Twitter crisis", "Twitter" AND "risk", "Twitter" AND "decision making". Additionally, keywords such as "social media" AND "swine flu", "social media" AND "ebola", and "social media" AND "healthcare" were used. Articles included in this review reported original research on Twitter for health, early warning detection, and an analysis of health-related research on Twitter. Articles written in English were selected.

Searches were also conducted in order to identify literature on social media and information theory and included "social media" AND "information theory", "social media" AND "information needs", "social media" AND "information seeking", "social media" AND "information behaviour". For literature on Twitter these search queries were adapted to "Twitter" AND "information theory", "Twitter" AND "information needs", "Twitter" AND "information seeking", "Twitter" AND "information behaviour".

Twitter was also used in this study for seeking research articles either by use of the platform, or by targeting specific hashtags, such as #NSMNSS, which stands for *New Social Media New Social Science* and is used by social media researchers.

2.3 Structure of Literature Review

Section 2.4 will outline information theory such as information, information needs, information seeking, and section 2.5 will outline a number of Social Cognition models and will provide a justification for the selection of the Health Belief Model for use in this thesis. It is important to understand the concepts of information, information needs, and information seeking as Twitter is a platform centred on the sharing of information via text, image, or video updates. Moreover, these concepts and theories will be later applied to the results of the study in Chapter 4 Swine Flu, Chapter 5 Ebola, and the smaller empirical chapter on Chapter 6 Zika.

Section 2.7 will outline the use of the Internet as a source of health information in a historical context, and will then go on to outline the social media platform Twitter (section 2.7.1). Section 2.8 will outline general health related research on Twitter, section 2.9 will outline how Twitter has been utilised for research on infectious disease outbreaks, section 2.10 will then outline research which has examined crisis communication using data from social media. These studies will be brought together in a comparator table (section 2.11). Section 2.12 will provide a synthesis of literature, and section 2.13 will provide an outline of current knowledge gaps, and how this present study seeks to address these.

2.4 Information and Related Concepts

It is important to consider concepts around information, information needs, and information seeking because Twitter is a platform which is centred around the sharing of information via text, images and videos. Section 2.4.1 will provide definitions of information which are relevant to this study, section 2.4.2 will outline the concept of information, section 2.4.3 will outline concepts around information needs, section 2.4 will outline the concept of information behaviour, and section 2.4.5 will outline information seeking.

2.4.1 Definitions of Information Relevant to this Study

When a topic of study is selected for further study, it becomes important for researchers to outline that which is under study as well as related concepts a term known as 'explication' (Case, 2012). Chaffee (1991) on the topic of explication has noted that it is a scholarly procedure, which will link theoretical concepts to research. Section 2.4.2 to 2.4.5 are related to the idea of explication, and will review and provide an overview of some of the terms that are relevant to this study such as the terms information, information needs, information seeking, and importantly, how these concepts relate to Twitter.

2.4.2 Information

As the concepts of information, information needs and information seeking will be used throughout this thesis, these are outlined in this chapter.

The word 'information' was first used between 1372 and 1386 in one of the poet Chaucer's tales (Schement, 1993; Case 2012). However, some have argued that the origin of the word 'information' may go as far back as the Latin and Greek ages (Capurro and Hjorland, 2002).

Information can come in a number of different forms, and it can have a number of meanings (Floridi, 2010). Case (2012) argued that we often use the word 'information' without thinking about its definition, because its meaning is intuitively easy to grasp, but there is no unique definition. Moreover, he noted that, even though people have been using the word information for so many years, researchers have still not come to an agreement on what the word really means. He noted that the issue with studying abstract concepts is that there will always be difficulty in defining them. This is because words can be vague and similar words can have a number of different interpretations. He provided the example of the noun 'port 'which can refer to 'fortified wine', the 'left side of a ship' and 'a gateway for opening passage'. However, he noted that the word information is even more difficult to define than port because with information there are a number of intersecting concepts that people may refer to. He argued that the *Oxford English Dictionary* may be adequate in defining the word 'information' (Case, 2012):

> "(1) The action of informing. The action of telling or fact of being told of something. (2) That of which one is apprised or told; intelligence, news" (p.58)

He noted that the definition does, at least, tell that the word information might be used to specify: 'i) a process, i.e. informing ii) or a kind of message, i.e. news' (p.48). Case (2012) adopted the following definition of information:

> "Information is whatever appears significant to a human being, whether originating from an external environment or a (psychologically) internal world" (p.87)

A review by Schement and Ruben (1993) found that there were at least over 20 definitions of information in the literature, showing how there is no unanimously accepted definition of the concept. This study will adopt the definition by Case (2012) provided above because it captures an important factor concerning of what information consists: were it not significant, it would not be of interest to humans who come across it.

The concept of information is relevant to this study as Twitter is a platform based on the premise of sharing information via text, image or video. Moreover, the information that is shared on Twitter has the potential to influence the behaviour of users. The definition by Case (2012) above is that an essential aspect of information is that it appears significant to humans and has some form of impact on the conscious mind. This definition seems appropriate in the context of information on Twitter. This is because a Twitter user, for example, might see that there had been flooding in area they were going to visit (i.e. this information would be significant to them), and this may cause them to alter their behaviour.

This section has outlined the concept of information, the next section will outline the concept of information needs, which is a related concept applicable to Twitter.

2.4.3 Information Needs

Case (2012) noted that other scholars (Wilson, 1981; Poole, 1985) argued that the idea of an information need is not a realistic concept, because most information needs could be classed as general needs, and they are not observable. They noted that people might have a need, for example to know how much an item such as bread costs; however, this would be driven by hunger, which is a basic human need. Nicholas and Herman (2010) defined information needs in two ways. Firstly, they suggested that information needs arise when a person requires information that can help with a task or research and the person recognises this (Line, 1974; Nicholas and Herman, 2010). Secondly, information needs are the result of people finding out that they have a gap in their knowledge and wish to resolve this (Belkin and Vickery, 1989; Nicholas and Herman, 2010).

Moreover, people may not have an information need, but rather when they experience a problem or carry out a particular task this may result in a cognitive need that could be met with the right information (Nicholas and Herman, 2010; Nicholas, 2003). Nicholas and Herman (2010) noted that information needs may arise to meet one of three basic human needs, which are:

1. Physiological needs, for example, the need for food, or shelter.
2. Psychological needs, for example, the need for domination or security.
3. Cognitive needs, for example, the need to plan or to learn a skill.

The needs above are related to Maslow's hierarchy of needs (Maslow, 1987). However, people's information needs may go unmet. This may be because, due to reasons specific to the person, they do not seek to meet the need (Nicholas, 2003; Nicholas and Herman, 2010; Case, 2012). Moreover, people may not know what their information needs are, or whether they have an information gap. This is because they may not know that there is information that could be of help to them. Only when exposed to relevant information the need is recognised by people and this is called this a 'dormant', or unrecognised, need (Nicholas and Herman, 2010; Case, 2012).

Nicholas and Herman, (2010) provided the example of a person overhearing a conversation related to a vegetable, known to contain harmful chemicals; if the person is fond of this particular vegetable they would listen with keen interest. The person did not have an information need, but they still obtained a piece of information. A further example for health information could relate to someone reading a newspaper and finding out about a risk-factor for cancer. They noted that people may be aware of their information needs, that is to say, that their needs are not dormant or unrecognised, but do nothing to meet them, as they are either unable to do this, or decide not to. They posited that even at the best of times information seekers tend to be 'satisficers' (satisficing being the conjoining of the words of 'sufficing' and 'satisfying') which was first defined by Simon (1956). 'Satisficing' means that they stop seeking information after locating material that is good enough (Savolainen, 2007; Nicholas and Herman, 2010), in order to juggle the need for comprehensive information with the constraint, for instance, of time or other resources, that may be placed on them.

Moreover, in today's internet-based world, information is generated in ever increasing volumes, and people are connected to information sources of unparalleled power and reach (Nicholas and Herman, 2010; Nicholas, 2003). Nicholas and Herman claimed that the popularity of the Internet stems from its unlimited potential to uncover dormant information needs in a searcher, and to solve recognised information problems expediently. However, they noted that turning to the Internet without a particular purpose in mind, frequently means relying on a 'happy accident which is also known as serendipity' (Foster and Ford, 2013).

The coverage of breaking news and emerging news stories via Twitter is well known (Bruns and Highfield, 2012; Vis, 2013). Therefore, users of Twitter may log into the platform (information seeking, outlined in section 2.7) if they have an information need, for example, if a person knew there was an emerging news story, possibly associated with a trending hashtag on Twitter, and they wanted to know what was happening where (information need). Research has shown that Twitter has the potential to bring value to those working within news media in the form of additional event coverage, as Twitter can report on events that are delivered by those who provide news (Petrovic, Osborne, Mccreadie, Macdonald, and Ounis, 2013). Therefore, as emerging news stories may now be reported on Twitter first (for example, citizens may record and upload videos and pictures of crisis events to Twitter, or bystanders may upload information to Twitter) users may now turn to Twitter if they have an information need (Zhao and Mei, 2013). With regard to Twitter, users may also log on with dormant information needs, i.e. when a person does not know what they are looking for. For example, a Twitter user may see that there had been a sale in a supermarket in close proximity to them, and this information may be useful to them. The person did not have an initially specified information need; nevertheless, it was satisfied. The following section will examine the concept of information behaviour.

2.4.4 Information Behaviour

Information behaviour relates to unexpected encounters with information, includes some elements of information seeking, and occurrences when people might actively avoid information (Case and Given, 2016). Furthermore, this term also includes “the broader context of how individuals ‘deal with ‘information in their lives, so accounts for situation, time, affect, culture, geography, and other contextual elements…” (Case and Given, 2016.p.6). Ford (2015) provided the example of research into the information behaviour of people who may have been affected by HIV/AIDS. There might be cases when relevant, timely, and up–to-date information on HIV/AIDs is available, but there will be other factors that influence whether people will be able to locate and consume this information. One of these factors is that some people suffering from HIV/AIDS may actively attempt to avoid information, and even hide or destroy it (Ford, 2015). It could be the case that people affected by HIV/AIDS may not be ready to accept such information out of denial, shock, and/or fear. Moreover, Ford (2015) noted that there might be cases where a public library might have information on HIV/AIDs, but a person might not be willing to be seen to be accessing such information publically out of fear of experiencing stigma and discrimination associated with the disease. In the context of health information on the Internet, some individuals who might not want access certain information, for example, information on a sexually transmitted disease, on shared computers such as in a library because of experiencing stigma. On Twitter, there may be certain topics that users do not want to be seen to be engaging with, for instance, advice on treating a sexually transmitted infection.

This section examined the behaviours associated with interacting with information, and the next section will outline the concept of information seeking which is another important concept.

2.4.5 Information Seeking

Belkin and Vickery (1989) noted that it is difficult to observe information needs as they occur inside people’s heads and can only be indirectly observed, and wrote that:

> "Less tractable is the issue of why people look for information at all; that is, what is the status of the concept or category of *information need?...*[I]s there such thing as a need for information, which can be considered on its own … or is information-seeking behaviour contingent upon the desire to satisfy other types of needs, or to resolve situations which are not in themselves information-dependent?" (p.6)

Case (2012) noted that this leads to information seeking and suggested that it may seem illogical that there has been a lack of research on information seeking and how it has been used, and noted that, when people study information seeking, they do not define the concept and it is assumed that it is something people do in response to having a need for information. He also suggested that only a handful of research defines information seeking, and it is often then defined as a way "discovering patterns or filling in gaps in patterns previously defined" (p.90).

Case (2012) provided a definition of information seeking from Zerbinos (1990):

> "Information seeking takes place when a person has knowledge stored in long term memory that precipitates an interest in related information as well as the motivation to acquire it. It can also take place when a person recognizes a gap in their knowledge that may motivate that person to acquire new information." (p. 922)

Case (2012) noted that information seeking may be a concept that is taken for granted. However, he noted that the term is the most used, but its current definition is restricted to behaviour which is active and intentional. This limits its uses to research looking at how people use information. There are some researchers who have examined information seeking in the context of electronic environments. Marchionini (1997), for instance, has noted that information seeking can be said to be an interactive process because it is based on the drive of the person seeking information, feedback from the environment, and decisions from the user based on the feedback from the environment.

Users on Twitter may use the platform for the purposes of actively seeking information because it may host information related to emerging news stories before any mainstream media outlet. This may lead to people using Twitter to *seek* information, because it is not yet available on other media, and information has the potential to spread very fast on Twitter. Zhao and Mei (2013) studied Twitter to examine types of information Twitter users would seek and set out requirements for when a tweet would convey an information need. They noted that a tweet would meet an information need if a user requested factual knowledge, an opinion, recommendation, or a confirmation of information. Information needs on Twitter were found to be based on real-world events, and altered with worldly events. It was also found that the questions people would ask on Twitter were different to the content tweeted on the platform.

2.5 Social Cognition Models Relevant to this Study

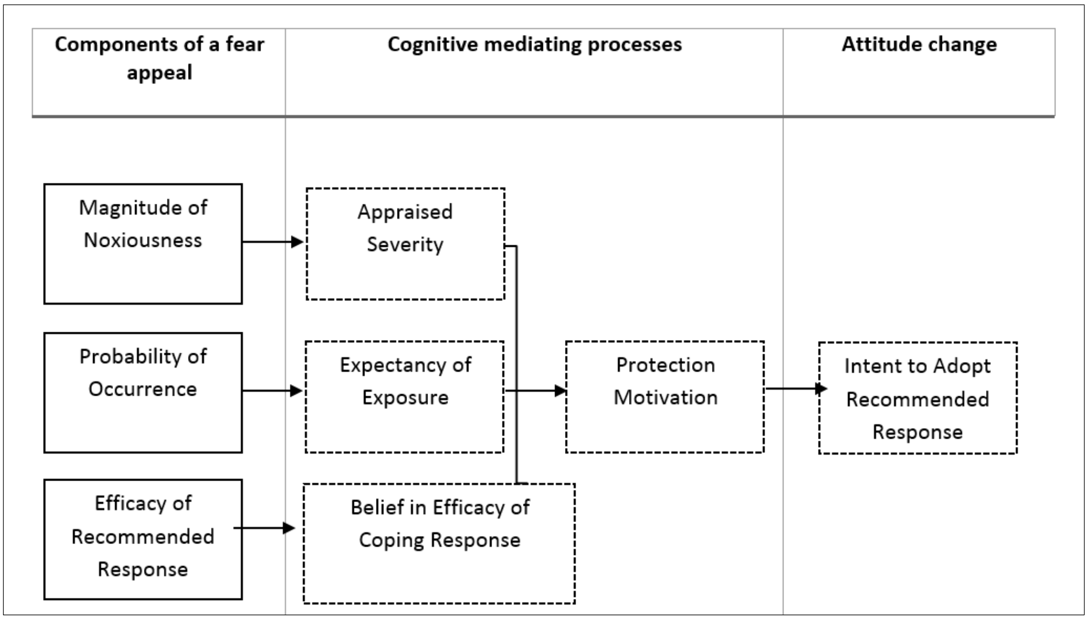
There are a number of theories relevant to the study described in this thesis that branch under the umbrella term of 'social cognitive models' which argue that mental reasoning, such as the beliefs and attitudes of a person, are factors influencing how a person behaves (Sutton, 2002). This section will provide an overview of a selection of these models/theories that are particularly relevant to this thesis, i.e. the Protection Motivation Theory (PMT), the Self-Efficacy Theory (SET), and the Health Belief Model (HBM).

2.5.1 Protection Motivation Theory

This section provides an overview of the protection motivation theory. Response-efficacy is the belief that undertaking a particular action would be effective in reducing a potential threat. Perceived self-efficacy is the belief that a person is able to carry out a behaviour that would be able to reduce a potential threat (Bandura, 1986). Thus, under the Protection Motivation Theory (PMT), a person would be motivated to avoid a potential negative threat and protect themselves if: 1) there was a chance they could get a disease, for example (probability of occurrence), 2) their risk was increased (expectancy of exposure), 3) the person believed a behaviour is effective, e.g. washing their hands to avoid swine flu (belief in efficacy of coping response) and 4) they thought they could do it, e.g. they have access to soap and water (protection motivation) (Sutton, 2002).

Figure 2-1 is a diagram of the protection motivation theory.

Figure 2-1 Diagram of protection motivation theory (PMT)

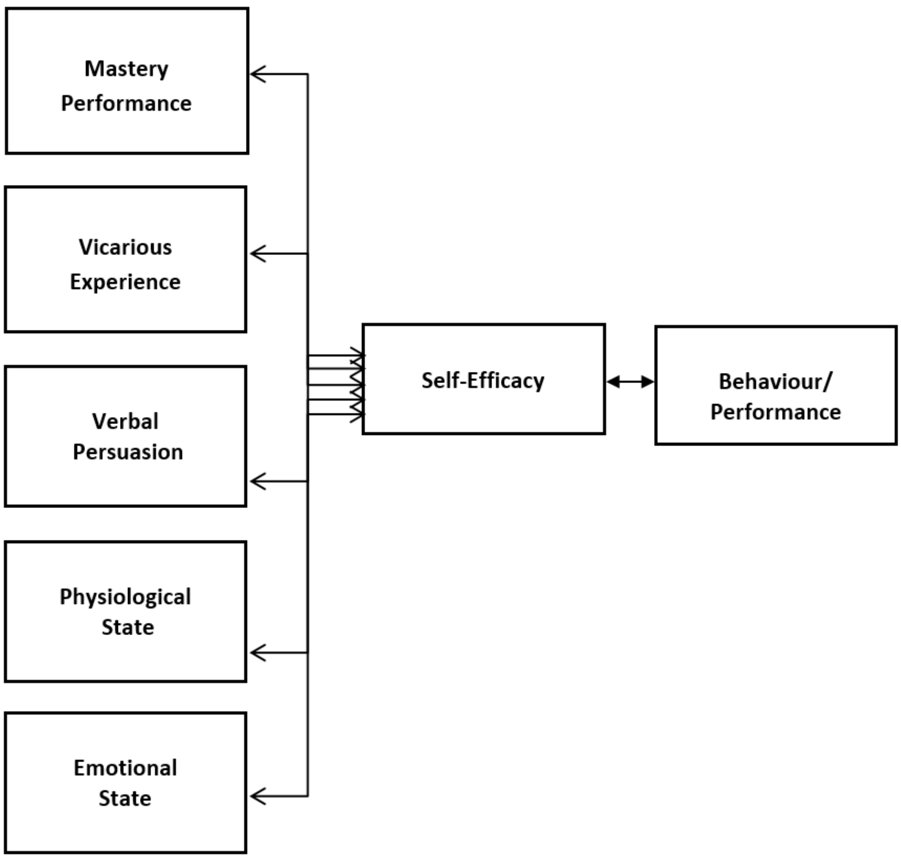


Recreated in Microsoft Word from: Rogers (1975).

2.5.2 Self-efficacy Theory (SET)

The Self-efficacy Theory (SET) is another social cognition model and is derived from Bandura's (1986) Social Cognitive Theory (Sutton, 2002). This model suggests that the two central factors which determine a person's behaviour relate to self-efficacy and outcome expectancies.

Figure 2-2 Diagram of self-efficacy theory (SET)



Adapted from Bandura (1997).

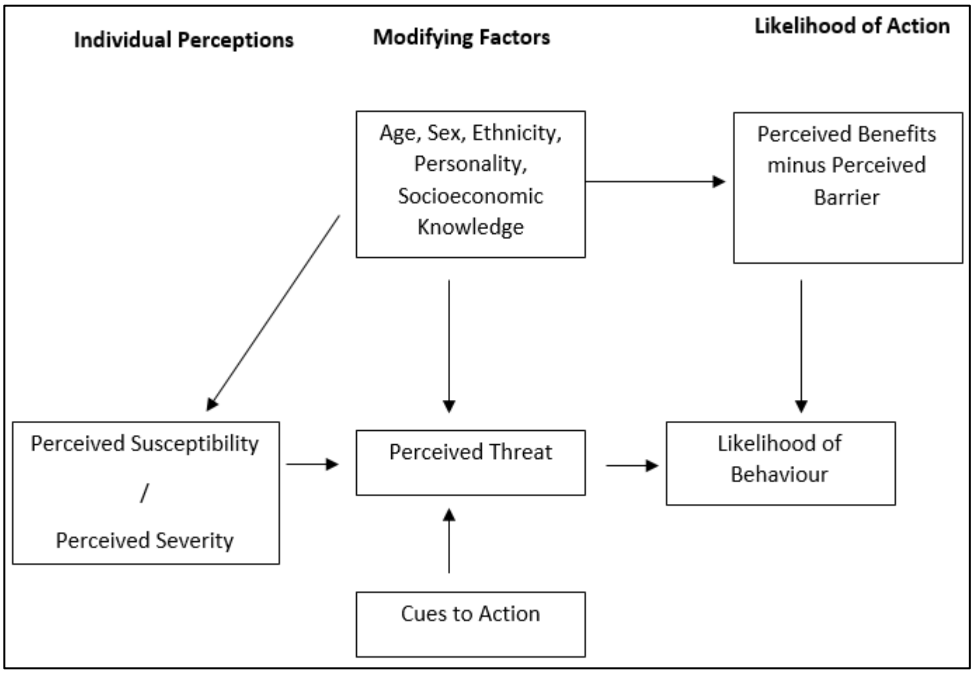
Self-efficacy is the belief that a person is able to adopt a particular behaviour; for example, regularly washing hands to avoid getting a disease or virus. As shown in Figure 2-2, according to SET, a number of factors such as 'mastery performance', 'vicarious experience', 'verbal persuasion', 'physiological state;, and 'emotional state' will have an effect on whether we perform a certain behaviour (Hayden, 2014). Outcome expectancy refers to what will happen if a particular behaviour is undertaken; for example whether there will be positive or negative consequence. In the case of a person washing their hands regularly, there will be the potential for a positive outcome as the risk of obtaining a virus or a disease would potentially decrease.

2.5.3 Health Belief Model (HBM)

The Health Belief Model (HBM) was formulated in the 1950s by a number of social psychologists who sought to find out why people were not using certain services such as immunisation (Becker, 1974), that is, when a person receives a vaccine in order to become immune to an infectious disease. Although developed in the 1950s, the model is still utilised to this day (e.g. Sutton, 2002; Guillaume, 2006; Hayden, 2014). The HBM has four main constructs. The first two constructs are related to the type of disease under study (perceived susceptibility and perceived seriousness), and the second two are related to the actions a person can take in order to reduce their chance of getting the disease (perceived benefits and perceived barriers) (Sutton, 2002). 'Perceived susceptibility' is based on a person's risk of obtaining a particular disease if they do not alter their behaviour, e.g. common cold (potentially high) and malaria (low in many countries). 'Perceived seriousness' is based on how the person thinks about the severity of a disease. For example, the common cold may have a low perceived severity, whereas an infectious disease such as malaria may have a very high perceived severity. The term 'perceived benefits' is based on the potential benefits of undertaking behavioural changes in order to reduce the risk of a disease, for example, a person who begins to wash their hands regularly during the outbreak of an infectious disease. The 'perceived barriers' construct is related to potential barriers of undertaking behavioural change in order to avoid a potential health threat (Becker, 1974; Sutton, 2002). For example, a government may recommend against travelling to a specific geographical region due to an outbreak of an infectious disease. However, if a person has family in the region, this may be a potential barrier for them. In the HBM, these different constructs are combined in order to weigh-up whether a person is likely to undertake a particular behaviour. According to the HBM, therefore, if a disease were to have a high severity, high susceptibility, high benefits, and low barriers then there would be a high chance that a person would undertake a recommended behaviour. In the context of cardiovascular disease, health authorities may recommend that citizens engage with behaviours that would decrease the risk of developing the disease such as walking regularly. Undertaking these actions would lead to high benefits, and as cardiovascular disease has a high severity and if recommended behaviours seem to include little or no barriers, according to the HBM, most people may be likely to follow these recommended behaviours. However, one of the limitations of the HBM is that it does not take into account the complex processes that underpin human decision-making and may attempt to simplify human behaviour, which may be more complex than the model.

The Health Belief Model (HBM) was developed in order to better understand and attempt to predict certain health-related behaviours, such as the subscription to health services (Becker, 1974; Janz, Nancy, Marshall, and Becker, 1984). HBM is frequently used in health promotion as well as health education (Glanz, Rimer, and Lewis, 2002; Sutton, 2002; National Cancer Institute [NCI], 2003; Hayden, 2014). The model was first developed as a way of explaining why certain medical programmes provided by the United States Public Health Service were not as successful as others (Hochbaum, 1958; Hayden, 2014). Figure 2-3 below, is a diagrammatic representation of the Health Belief Model.

Figure 2-3 Health Belief Model (HBM)



Recreated in Microsoft Word from: Stretcher, V., and Rosenstock I.M. (1997). The Health Belief Model in Glanz K., Lewis F.M, and Rimer B.K., (Eds.). Health Behaviour and Health Education: Theory, Research and Practice. San Francisco: Jossey-Bass.

As indicated in Figure 2-3 above, one of the key concepts of the HBM is that health behaviour can be driven by personal beliefs towards a particular disease and beliefs concerning the approaches that are available in order to reduce the likelihood of developing a disease (Hochbaum, 1958; Hayden, 2014).

The perceived seriousness construct refers to an individual's belief about the seriousness or severity of a disease (Hayden, 2014). Perceived seriousness may be based upon medical information, or may arise from the beliefs a person holds with regard to the difficulties a disease may create, and the impact that these have on a person's day to day life (Hayden, 2014). For example, cancer may be perceived as serious due to its high mortality rate or due to beliefs a person may hold about it.

The perceived susceptibility construct is based on the idea that the greater the perceived risk, the greater the likelihood that a person will engage in behaviours that will decrease the risk

(Hayden, 2014). In other words, if people believe that they are at risk of contracting a disease they will be more likely to act in a way to prevent it from happening, e.g. use a condom to avoid HIV/AIDS infection. Moreover, if the perception of susceptibility is combined with seriousness, this results in a perceived threat (Stretcher and Rosenstock, 1997; Hayden, 2014). For example, if a person is obese, their susceptibility to heart disease may be high, and as the disease is very serious, this could result in a high perceived threat. However, it is important to note that this is a theory that has been purported, and behaviour change may or may not occur depending on the individual, the situation and a number of other factors.

An outbreak of bovine spongiform encephalitis (BSE), commonly known as ‘mad cow disease’, led to reduced consumption of meat. BSE affects the brain causing tiny holes that make it appear sponge like (Hayden, 2014). In humans, this can lead to ataxia (the lack of voluntary coordination of muscle movements), and can also lead to dementia (Stoppler, 2016). The HBM can be applied to BSE because the perception of threat was associated with the consumption of beef, and thus some people ceased to consume it (Weitkunat *et al.*, 2004).

Coe, Gatewood, Moczygemba, Goode and Beckner (2012) used the HBM to understand better why people would want to receive the H1N1 influenza vaccine. The participants in their study felt that they were at risk of H1N1, or worked with vulnerable people (children, the elderly, the sick) who had a higher perceived seriousness of illness if they contracted it, and so planned to take the vaccine. Thus, if members of the public are educated on the benefits of vaccines, then they may be more likely to opt in for a vaccination and this strategy could be used in future outbreaks to increase the rate of vaccinations and decrease contagion.

The construct of perceived benefit refers to the idea that people tend to adopt healthier behaviours when they believe that such behaviours will decrease the chances of developing a disease (Hayden, 2014). In the case of a disease such as obesity, this could translate to behaviours such as regularly walking or being careful with their diet. In the case of a disease such as cancer, this could translate to quitting smoking.

The construct of perceived barriers refers to the idea that, in order for a new behaviour to be adopted, a person would need to believe that the benefits of the new behaviour would outweigh the consequences of continuing the old behaviour (Centres for Disease Control and Prevention, 2004; Hayden, 2014). In the case of cancer, this quitting smoking is not perceived as so bad as developing cancer.

The four major constructs of perception can be modified by other variables such as culture, education level, past experiences, skill, and motivation (Hayden, 2014). There are also individual characteristics that influence personal perceptions. For example, if someone was treated for a melanoma they may have a heightened perception of susceptibility due to this past experience and may be more likely to avoid sun bathing in future.

Hayden (2014) labelled cues to action as events, people or things that lead people to change their behaviour. Examples include illness of a family member, media reports (Graham, 2002), mass media campaigns, advice from other people, reminder postcards from health care providers (Ali, 2002). For example, when people hear radio adverts about instructions on how to safely handle raw food, these are known as *cues to action* that are associated with safer food-handling behaviours (Hanson and Benedict, 2002) to reduce chances of food poisoning.

Self-efficacy is the belief in a person's ability to do something (Bandura, 1997; Hayden, 2014), because people generally do not try new things unless they feel that they can do it (Hayden, 2014). For example, if a person thinks a new behaviour is useful (perceived benefit), but they feel they are not capable of doing it (perceived barrier), there is a chance that this behaviour will not be attempted (Hayden, 2014).

In 2003, when there was an outbreak of the severe acute respiratory syndrome (SARS) disease, it was perceived as a serious disease. People may have thought that their susceptibility to it was high, i.e. that there were benefits of adopting new behaviours of avoiding it, and that the advantages of adopting a new behaviour outweighed any disadvantages of adopting the behaviour. This means that for SARS, people may be likely to adapt their behaviour. Some may use Twitter to view a news story, or deliberately visit the platform to seek information due to a gap in knowledge (as explained in Section 2.7). Knowing more about a disease especially in relation, for example to its geographical spread, may cause people to change their behaviour.

It has been suggested that Twitter may be a good platform for social support and to motivating people to alter unhealthy behaviours (Nelson and Staggers, 2013). In applying the HBM model when disseminating information via Twitter, it is possible to understand both why users may uptake certain behaviours (i.e. avoiding certain areas) and why they may not modify other behaviours. In the discussion of results (section 7.5), the constructs of the Health Belief Model will be further applied within this thesis. This is because the HBM is more suitable for infectious disease outbreaks, and was initially established in order to better understand why people would not opt in for immunisation against such diseases. HBM is based on a number of constructs such as individual perceptions, modifying factors, and the likelihood of action, and is suitable for providing insights into how Twitter users behave during an infectious disease outbreak.

2.6 Sociological Concepts

There are a number of sociological concepts such as the 'moral panic' and 'outbreak narrative' which will be applied to the results of the study, and are outlined next.

2.6.1 Moral Panic

The concept of moral panic is a well-established sociological notion first developed in 1972 (Cohen, 2002). The concept is summarised in the passage below:

> "Societies appear to be subject, every now and then, to periods of moral panic. A condition, episode, person or group of persons emerges to become defined as a threat to societal values and interests; its nature is presented in a stylized and stereotypical fashion by the mass media; the moral barricades are manned by editors, bishops, politicians and other right-thinking people; socially accredited experts pronounce their diagnoses and solutions; ways of coping are evolved or (more often) resorted to; the condition then disappears, submerges or deteriorates and becomes more visible. Sometimes the object of the panic is quite novel and at other times it is something which has been in existence long enough, but suddenly appears in the limelight. Sometimes the panic passes over and is forgotten, except in folklore and collective memory; at other times it has more serious and long-lasting repercussions and might produce such changes . . . in legal and social policy or even in the way the society conceives itself." (Cohen, 2002: p.1)

Garland (2008) outlined a number of examples of potential moral panics which include: the witch-craze that took place in the 16th and 17th centuries, the war on drugs, and threats from terrorism. Moral panics could turn out to be insignificant later on, but at the time they occur they will generate a real sense of panic whilst they are happening. A typical feature of moral panic is an overblown reaction from different parts of society. This concept could potentially be relevant to this study because infectious disease outbreaks are further examples of moral panics.

2.6.2 Outbreak Narrative

In *Cinematic Prophylaxis*, Ostherr (2005) provided an overview of how historical information about disease transmission can affect the public's understanding of a disease. Specifically, Ostherr (2005) noted that 'cinematic prophylaxis' is the cross discipline study of public health and Hollywood outbreak narratives. In outbreak narratives shown on television, members of the public see facemasks, decontamination procedures, and images of slaughtered animals, hence modifying how they might think about public health at times of real outbreaks. Therefore, when an outbreak does occur people might draw on these narratives which are shown across Hollywood films, video games, and in news coverage. Certain media narratives may be factual whereas others will be based on fiction, for instance, stories of zombies and the potential of a zombie apocalypse are examples of fictionalised narratives. Ostherr (2005) noted that audio-visual materials are metaphors which provide information on a disease and how it might spread. During an outbreak of a deadly disease, media organisations might want to respond immediately and there might not be time for them to consider the most appropriate methods of disseminating information during an outbreak. Outbreak narratives can be found across film, novels, popular culture, and via historical events (Wald, 2007). This section has explored sociological concepts which will be applied to the results of this study and the next section examines the Internet as a source of health information.

2.7 Internet as a Source of Health Information

In the late 20th and early 21st century, the Internet has emerged as a source for accessing medical information (Diaz, Griffith, Reinert, Friedmann, and Moulton, 2002; Morahan-Martin, 2004). Early research into health information on the Internet noted that it had the potential to influence the decision making of members of the public. An early study on health information on the Internet found that up to 70% of Internet users reported that the content on the Web influenced their decision on the treatments they would receive (Fox and Rainie, 2000). People who require health information from the Web can access it any time of the day with complete anonymity (Morahan-Martin, 2004). During emerging disease outbreaks, therefore, people may navigate to World Wide Web in order to locate information because the information is private and instantly available. In the 2003 SARS outbreak, the disease emerged as among the top 20 most frequently searched for terms over a two-month period (Yahoo! Buzz Index, 2003; Morahan-Martin, 2004). Moreover, research has found that when people go online for health information people will perceive the information to be trustworthy (Fogel, Albert, and Schnabel, 2002). Tanis (2008) noted that in addition to members of the public utilising the Web to seek information that people have also made extensive use of online forums (originally formed to share and discuss news) to seek and/or share health information. This could involve the exchange of information, discussing problems, posting stories, and being empathetic to one another (Tanis, 2008).

Throughout the history of the Internet, people have utilised the online world to seek, share, and search for information by using the World Wide Web, and online forums (Diaz, Griffith, Reinert, Friedmann, and Moulton, 2002; Morahan-Martin, 2004; Tanis, 2008). In the early years of the 21st century, Web 2.0 and social media platforms emerged, which were originally intended for people to connect with each other socially and share news stories, images, and videos with one another (Kaplan and Haenlein, 2010). The use of these platforms emerged for the sharing of health information, and for health related communities to form on the platforms (Nelson, and Staggers, 2013). An emerging platform for studying health topics for public health research is Twitter. As noted in the paragraph above, during the SARS outbreak, there was an increase in people seeking information on the disease. Twitter has made it easier to follow emerging events and health scares and for members of the public to share their opinions. Section 2.6.1, next, will provide an overview of Twitter, and section 2.7 will provide an overview of the types of research that have been conducted on the platform.

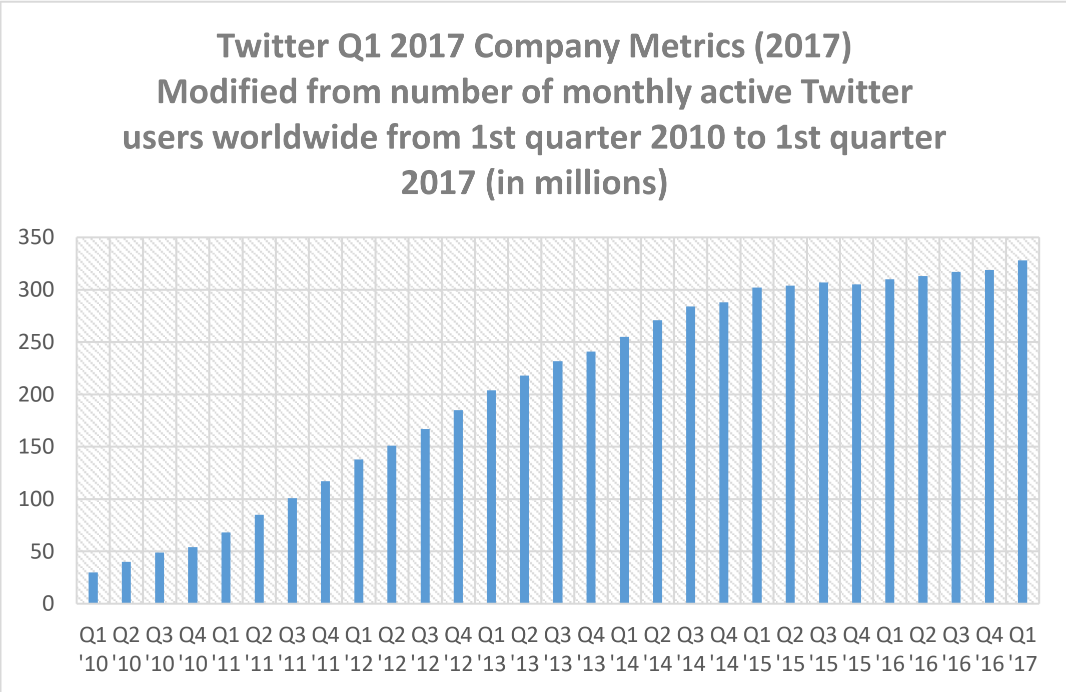
2.7.1 Twitter

Twitter was launched in 2006 (Weller *et al.*, 2014), and is a microblogging service which serves as a platform for information flow where users can post updates and subscribe to other users, known as 'following', in order to receive updates or microblogs from other users (Purohit *et al.*, 2013). Twitter does not disclose demographic data (Sloan, 2017) on those who use Twitter; however, a report by Pew Research (2016) from the US reported the following usage percentages for all online adults:

- 36% of 18 to 29 year olds
- 23% of 30 to 49 year olds
- 21% of 50 to 64 year olds
- 10% of 65+ year olds

The report also found that 29% of adults who graduated from college use Twitter, and 25% of users have some form of college education with 20% of users holding a diploma from higher school. The report also found that 21% of all Americans use Twitter, and 79% of all Twitter accounts are from people outside the US. During the first quarter of 2017, Twitter had an estimated 328 million monthly active users (Twitter Q1 2017 Company Metrics, 2017). Twitter published company notes indicating that it has one billion unique visits monthly, and that 82% of active users are on mobile devices (About Twitter, n.d). Figure 2-4 below displays how the total number of monthly active users has increased on Twitter from 2010 to 2017. This is important for this present study as there is a large increase in the number of users from 2009 to 2014.

Figure 2-4 Twitter numbers of monthly active users from 2010 to 2017



This study will highlight features of Twitter from Purohit *et al.* (2013), who provided an overview in a research paper. Tweets are short messages also known as microblogs or posts which contain information from users on their updates, actives, and users may also share news articles. Users can also exchange tweets privately through *Direct Messages*, and these tweets are only viewable to the sender and recipient of the messages. Twitter users may link to external sources in their tweets such as websites. In the earlier years of Twitter users would utilise services such as http://bit.ly/ to shorten the URLs length in order to avoid lengthy tweets. A hashtag begins with the '#' symbol followed by a word (e.g. #EbolaOutbreakAlert).

Hashtags are a platform convention which can be created by users and are used to share information related to a single topic. When there is an emerging news story, the discussion around it will often be based on a hashtag. The 'reply' feature is a method used to reply to a tweet. The 'retweet' function will share i.e., forward a tweet from a user to their followers. There were also a number of features which were added to Twitter after the study by Purohit *et al.* (2013) was published. That is, it is now possible to retweet another user by quoting them e.g., users tweet '[Original tweet]' as '@userhandle I concour [@userhandle1 today is amazing day]'. A term known as 'trending' on Twitter refers to a topic (a keyword or hashtag) that is popular at a specific time. Although this is not an exhaustive list, it provides an overview of some of the central features of Twitter which were deemed relevant to this study.

2.8 Health Related Research on Twitter

The previous section outlined the Internet as a source of health information and then provided an overview of Twitter. The aim of this section is to provide an overview of studies that have investigated Twitter for general health research.

Twitter has provided researchers with the opportunity to gather and analyse health-related Twitter data (Sinnenberg, 2017). There are studies that examine Twitter for a specific health condition, gather data, and analyse data using a range of quantitative and qualitative research methods. In addition, there are studies that have examined how Twitter users seek and share health-related information on Twitter, or how health organisations disseminate health information.

A broad range of health topics can be researched on Twitter. One study by Scanfeld, Scanfeld and Larson (2010) analysed 1,000 tweets which mentioned antibiotics using content analysis in order to explore possible evidence of their misunderstanding or misuse. The authors suggested that social media websites allow users to share health information, and could be used to identify and understand the misuse of antibiotics and promote positive behavioural changes. They noted that, in August 2009, only 94% of all tweets were publically available, and as such not all the relevant data could be gathered from Twitter. The period of data collection also coincided with the H1N1 outbreak, which could have influenced on the types of tweets users were sending. They concluded by arguing that it is important for health authorities to have a basic grasp of social media such as Twitter. This is because Twitter could be used to gather real time health data and could be used to identify and understand the misuse of antibiotics.

Researchers have also examined discussions surrounding cardiac arrest on Twitter. Bosley, Zhao, Hill, and Shofer (2012) investigated how users of Twitter sought and shared information on this topic. Cardiac arrest is a time-sensitive condition where initial treatment is dependent on the public awareness and response available at the time. They noted that large volumes of tweets can be filtered to identify specific public knowledge and the quality of information seeking and sharing about cardiac arrest. As a result, they suggested that Twitter provides access data that would otherwise be out of reach for health research. The limitations of the study were that Twitter users are not representative of the whole population of people most at risk of a cardiac arrest in the United States nor in other countries, and that the keywords used may not have covered all the words people commonly use to talk about cardiac arrest. The limitation raised in the previous study is applicable to all studies that gather tweets on a single topic. Users might tweet about a topic without mentioning the keyword and/or hashtag under study, making it difficult to gather all data on a certain topic using keywords and hashtags.

Myslin, Zhu, Chapman, Conway (2013) examined the content and sentiment related to new tobacco-related products such as hookah and electronic cigarettes. The study gathered 7,362 tweets over a 15-day interval spread across the time period from December 2011 to July 2012. The study made use of machine learning in order to automatically detect tweets that were relevant, as well as to eliminate tweets that were not. The study found that the themes that occurred with most frequency related to 'hookah', 'cessation', and 'pleasure'. The study concluded by suggesting that it was possible to use Twitter for novel insights related to tobacco surveillance.

Tweets related to dementia have also been studied. Robilard *et al.* (2013) investigated how information on age-related diseases such as dementia is disseminated on Twitter. They reported three key findings; firstly, that the majority of tweets referred to other websites such health information or news websites, rather than the sharing of personal information; secondly, they reported that the majority of tweets originated from users who identified themselves as health professionals, or who were associated with a health or news website; and thirdly, they reported that most discussion was formulated around recent studies on prediction and risk of dementia. The limitations of the study included possible selection bias, and the difficulty of assessing whether the sample of tweets was representative of the national offline population.

Researchers have also studied the association between suicide and suicide-related tweets. Sueki (2014) conducted an online cross-sectional survey of 20-year old Twitter users in Japan. The aim was to identify an association between suicide-related tweets and suicide behaviour. Sueki (2014) employed logistic regression analysis and found that tweeting 'want to commit suicide' was a predictor of future suicide behaviour. Sueki (2014) concluded that Twitter logs or logs from social media platforms could be a useful predictor of suicide behaviour in young adults. However, the study was limited as it focused on Twitter users based in Japan who were registered with one particular internet survey company. Moreover, tweeting 'want to commit suicide' may not be the only phrase that is used to signify that a person is considering attempting to commit suicide. A more comprehensive study would have employed multiple search queries.

Sexual risk behaviour communication using geolocation data has also been studied on Twitter. Young, Rivers, and Lewis (2014) investigated whether and how geo-located conversations on HIV risk behaviour could be extracted from Twitter, the content and occurrence of these conversations, and the feasibility of using HIV risk-related conversations on Twitter in order to detect HIV outcomes. It was found that there was a significant positive correlation ($p < 0.01$) between HIV-related tweets and the occurrence of HIV cases in the US. They suggested that using real time social media data to map the occurrence of HIV is feasible. In this study, data for HIV were only available from 2009, which could mean the HIV tweets could have come from regions where HIV rates were already high. Hence, the results could have occurred by chance, because if people already lived in an area where HIV is predominant they would be more inclined to tweet about it. The authors nevertheless suggested that this was a feasibility study to demonstrate a causal link between HIV tweets and HIV prevalence, and that this could be mapped inexpensively using a data source such as Twitter.

Researchers have also examined discussions surrounding childhood obesity on Twitter. Harris, Moreland-Russell, Tabak, Ruhr and Maier (2014) used NodeXL to retrieve tweets that were sent in June 2013 and which matched the hashtag #childhoodobesity, and coded these for content as well as actor type (i.e. Twitter user). NodeXL is an open source template which can be used to create network graphs by importing graph data, or by retrieving it through social media platforms (Hansen, Shneiderman, and Smith, 2010). The study found that individuals rather than organisations would use the account and that tweets were related to the individual behaviour of users. The study noted that there is the potential to disseminate relevant and credible evidence-based information to the diverse set of users on Twitter by using the #childhoodobseity hashtag and by focusing content on scientific evidence.

Kendra, Karki, Eickholt, and Gandy (2015) examined Twitter data in order to see whether it was possible to obtain information related to health topics from social media, and also to gain insight into how users were conversing about antibiotics. The study used 89 keywords related to antibiotics and collected 591,091 tweets over a three-month period (27th May to 11th September 2014). Many users were found discussing antibiotics and the most frequent topic of discussion was antibiotic resistance. Moreover, the study found that there were a number of national news events which would mention antibiotics and which would cause a sudden increase in the frequency of Twitter activity. The study also reported that a number of open source tools such as Hive, Flume, and Hadoop have the potential to gather and analyse tweets.

On Twitter, it is also possible to conduct analyses on whether certain Twitter users are influential. Cavazos-Rehg *et al.* (2015) set out to examine sentiments and topics discussed by influential users on Twitter related to marijuana. The study used the influence metric of the Klout score in order to measure social influence and identify influential Twitter users. The Klout score is an online algorithm which produces an influence score when a person connects their social media platforms to the service (Klout, n.d.). A total of 7,000 tweets were retrieved for users who were ranked among the top 25th percentile due to their Klout (n.d.) score as well as based on the number of followers that they had. The study did not publish the identity of the influential Twitter users. It was found that, among the sample of tweets, the pro-marijuana tweets outnumbered the anti-marijuana tweets by a factor >15. The study stated that the findings were important as they could be used to better inform both offline and online prevention efforts and people who may have a higher risk for being marijuana users. The limitations of the study were that tweets were retrieved over a one month period, and only a limited number of marijuana keywords were used.

Research has also examined diabetes discussions on Twitter. Liu, Mei, Hanauer, Zheng, and Lee, (2016) monitored the number of diabetes-related keywords over a two-year period using the Gardenhose API (a method of retrieving Twitter data as explained in Methodology Chapter 3 section 3.8), which provides a 10% sample of from Twitter. In the study, 29.6 billion tweets were retrieved from Twitter from 2013 and 2014, and it was found that 1,368,575 tweets within the dataset contained diabetes-based words and hashtags. The study confirmed that Twitter was a platform where members of the general public would tweet about diabetes. The study provided new insights on the rate at which people tweet about diabetes. However, no qualitative analysis techniques were applied to the data. Therefore, it is not known to what extent bots (i.e. automated accounts) and news stories had been tweeted and retweeted. As no content or thematic analysis was conducted, it is also not known what the key topics of discussion consisted of. The study by *Liu et al.* (2016) also discussed the concept of bots. Varol *et al.* (2017) estimated that 9% to 15% of all accounts on Twitter consist of bots. A bot is an account which is controlled by software and which may automatically generate content and interact with other Twitter accounts (Varol *et al.*, 2017).

The studies outlined above have highlighted the use of Twitter data gathered on a variety of health-related topics and analysed for content or examined how heath organisations disseminate information. The studies have reported that the benefit of using Twitter data is due to its accessibility and feasibility. Commonly reported limitations are that Twitter users are not representative of the general population, and that there might be possible selection bias when Twitter data are gathered. However, as Twitter data have allowed researchers to examine health topics that would otherwise have been out of reach, the reported advantages outweigh the limitations.

In summary, the studies outlined in this section utilised Twitter data to gain insight into a number of different health topics ranging from diabetes, marijuana, antibiotics, childhood obesity, sexual risk behaviours, suicide, dementia, tobacco-related products, cardiac arrest, and antibiotics. This highlights how Twitter data has been utilised to study a range of different health topics and that there is varied discussion on Twitter related to health. The next section provides an outline of studies conducted specifically on infectious disease outbreaks.

2.9 Infectious Diseases

The previous section provided an outline of studies which utilised Twitter data in research for general health topics. This section will outline studies that have specifically explored infectious disease studies and surveillance on Twitter.

The 2009 swine flu (H1N1/09) pandemic was the first influenza pandemic since 1968 which carried a Word Health Organisation (WHO) threat level of 6 (the highest). This gave rise to a large volume of studies examining public sentiments and early warning systems for the H1N1/09 outbreak (swine flu). In one of the earliest published studies examining Twitter in relation to pandemics, Chew and Eysenbach (2010) monitored mentions of swine flu during the 2009 pandemic on Twitter. The study sought to understand whether Twitter could be used to gather public views and opinions on a large scale. The study found that there was no prior methodology for sampling tweets, and they were unable to perform sample size estimates. They utilised a mixed method approach involving machine learning and content analysis. The research found that the tweets related to H1N1 consisted of sharing credible information and the opinions and experiences of users. The results suggested that Twitter was a rich source of information during the pandemic. Consequently, the authors recommended that health authorities "become aware of and respond to real or perceived concerns raised by the public" (Chew and Eysenbach, 2010, p.12). This was a significant piece of research, as it used Twitter as a primary data source for views and opinions related to the swine flu outbreak. They noted that Twitter is a valuable platform that health authorities should become monitor. The study was published when Twitter had far fewer registered users (in Q4 of 2010 the maximum user base of Twitter was 54 million monthly), and its usage and thus potential have since expanded.

Paul and Dredze (2011) noted that a variety of public health topics could be studied on Twitter and raise the point that great insight could be gained by aggregating millions of tweets and examining their locations and the languages using the tweets. They also noted that Twitter could be utilised to disseminate health information for the purposes of spreading health information in regards to crisis and risk communication. In the study by Paul and Dredze (2011) a model was applied to 1.5 million health-related tweets in order to better understand 12 aliments such as allergies, obesity, and insomnia. They were able to track aliments both over time and geography highlighting how public health organisations could utilise data from Twitter in place of more expensive surveillance methods.

In another study on swine flu on Twitter, Signorini, Segre and Polgreen (2011) examined public sentiments on Twitter related to the H1N1/09 (swine flu) outbreak of 2009 with the aim of tracking the spread of H1N1/09 by using data from Twitter. Traditional methods of disease surveillance consist of building a map of the geographical range of an infectious disease utilising data from literature, reports from the web, and certain databases (Hay, George, Moyes, and Brownstien, 2013). They noted that Twitter could be used both for tracking user's interests and concerns, but also for tracking and estimating disease activity one to two weeks faster than traditional methods. They suggested that Twitter could be used to measure public sentiment related to health topics and events.

Kostkova, Szomszor, and St. Louis (2014) also examined the swine flu outbreak of 2009 in order to demonstrate the efficiency of early warning systems. However, the study is distinct from that of Chew and Eysenbach (2010) outlined above, because rather than validate Twitter as a means to gather public views on a large scale Kostkova, Szomszor, and St. Louis (2014) sought to demonstrate the potential of early warning systems. An early warning system is one which attempts to predict the occurrence of adverse events. The years 2009 and 2010 saw an increased interest in the use of the Internet to estimate the levels of flu by aggregating search queries (e.g. from Google search engine). They noted that just as Web search queries could estimate the incidence of flu that social media had the potential for understanding public concerns and act as early warning systems. They suggested that as social media platforms (such as Twitter) are in the public domain and publically available, they can be utilised for this purpose. The study searched for the keyword 'flu' and retrieved 3 million tweets, and they found a correlation between Twitter and UK and US surveillance data. The authors also found that Twitter provided an excellent method of sampling large populations and for tracking and predicting infectious disease outbreaks. Building early warning systems using Twitter data could potentially face similar challenges that have been faced by Google flu trends (Lazer, Kennedy, King, Vespignani, 2014), which was an over-reporting of flu incidence due an unexpected rise in search queries. Lazer, Kennedy, King, and Vespignani (2014) identified problems that had occurred with Google flu trends and other predication based algorithms. They noted that Google flu trends was not a stable reflection of flu incidence because of the way Google were managing and overfitting the algorithm.

In a study specifically aimed at the Portuguese language, Santos and Matos (2014) used machine learning to investigate whether user-generated content could be employed to provide a predictive model to obtain instant feedback on the incidence levels of flu in Portugal. A Naïve Bayes classifier was trained to identify tweets which would mention flu-like symptoms. They suggested that, with search engine queries, it may be possible to predict and estimate influenza-based illnesses in Portugal. They noted that an advantage of using Twitter and web search queries compared to traditional (often weekly or monthly) reports is instant feedback. This might prove extremely important for early detection and can consequently reduce the impact of epidemic breakouts because it would be possible to allocate resources in a timely manner.

There is research suggesting that rumours and medical misinformation may have been present on Twitter at the height of the Ebola epidemic. For example, Jin *et al.* (2014) found that, in addition to news reports shared on Twitter, there were conspiracy theories, innuendos, and rumours circulating on the platform. Tweets were gathered from late September, and the study utilised computational methods such as a dynamic query expansion model to study the rumours. Rumours were found to focus on the Ebola vaccine only working on white people, Ebola patients rising from the dead, Ebola becoming airborne, and health officials injecting Ebola patients with the disease.

Oluwafemi, Elia, and Rolf (2014) published a letter to the editor to the British Medical Journal (BMJ) on a study they conducted on Twitter and Ebola. They searched Twitter for tweets that contained the keywords 'Ebola', 'prevention', and/or 'cure', which were sent from Guinea, Liberia, and/or Nigeria from the 1st to the 7th of September. These tweets were then grouped into those which were medically correct, and those that were medically incorrect. They found that rumours on Twitter were based on alleged cures for Ebola, such as the plant Ewedu, blood transfusions, and drinking and washing in salt water. They found that occasionally inaccurate information was corrected by Nigerian government agencies. The authors found that occasionally inaccurate information was corrected by Nigerian government agencies. However, they recommended that government authorities within West Africa should be using Twitter in order to spread correct information.

Kalyanam, Velupillai, Doan, Conway, and Lanckriet (2015) examined potential misinformation and rumours on Twitter related to the outbreak of Ebola in 2014. Microblogging platforms such as Twitter have been shown to be useful for spreading information across the globe within seconds. However, Kalyanam *et al.* (2015) noted that misinformation and rumours are also as likely to spread on Twitter and be perceived as legitimate sources of information.. From a public health perspective, the spread of misinformation can create unnecessary panic for the general public. The authors used a dataset of 47 million tweets obtained from Twitter's streaming API and used quantitative methods by using a machine learning classifier in order to assign tweets into two sets, credible or speculative. They found that the number of hashtags per tweet in the *credible* set was slightly higher than that of the *speculative* set.

A common finding identified among the two studies undertaken on tweets on the 2014 Ebola epidemic are that Twitter contained false information (Oluwafemi, Elia, and Rolf, 2014; Jin *et al.*, 2014; Kalyanam, Velupillai, Doan, Conway, and Lanckriet, 2015)

Medical misinformation related to the cures for Ebola has the potential to take lives as, for example, some media outlets reported the rumour that salt water can cure Ebola and led to a number of deaths (DNews, 2014). However, it was not possible to identify a more authoritative academic source to verify this news story. There is a lack of empirical research examining medical misinformation on Twitter. However, more generally, medical information on the Internet has been studied widely and fears over incorrect information have been raised from the very early years of the Internet (Eysenbach *et al.*, 1998), as well as more recent studies such as Mocanu, Rossi, Zhang, Karsai, and Quattrociocchi (2015).

On Twitter, it is possible to conduct research on specific keywords, hashtags and user handles. It is also possible to conduct research on Twitter chats where Twitter users may talk about a shared interest using a hashtag over an allocated time period. Scheinfeld, Bernhardt, Wilcox, and Suran (2015) applied textual analytics which parse and can extract information from text on tweets related to a live Twitter chat that was organised by the Centers for Disease Control and Prevention after the diagnosis of Ebola on US soil caused widespread panic. Their aim was to inform public health authorities on the types of information the general public may require during an infectious disease outbreak. Tweets were collected, sorted and analysed using SAS Text Miner 12.1 in order to provide insight into the major themes within the dataset. Based on the analysis of the Twitter chat, they found that the public were concerned with the symptoms of Ebola, the lifespan of the virus, disease transfer and contraction, travel safety, and how to protect one's body from Ebola. It is important to note that this study is distinct from other conducted on infectious diseases because it examined tweets around a specific Twitter chat rather than all tweets on the platform.

Odlum and Yoon (2015) suggested that Twitter had the potential to address challenges that had arisen from conducting surveillance during the 2014 outbreak of Ebola. The aim of their study was to outline how Twitter could be applied as a real-time tool to monitor the spread of information, for epidemic detection, and to use Twitter as a source of public views and attitudes. In total 42,236 tweets were retrieved from the time period of July 24th to August 1st 2014, and consisted of English language tweets which mentioned Ebola. The data was analysed using quantitative methods such as time series analysis and natural language processing. An increase in tweets was observed two to seven days before Ebola was officially announced to have occurred in Nigeria. The study was conducted retrospectively therefore this finding was not applied in practice. The authors identified a number of different discussions taking place at the same time on risk factors of Ebola, prevention, education, disease trends, and compassion. The study noted that the results were able to provide insight into the usefulness of Twitter to inform public health education, and for public health surveillance. However, the study was noted it was limited, as it only considered tweets in English and looked at a single social media platform (Twitter).

Towers *et al.* (2015) looked at the media coverage of Ebola in the US and explored whether this had affected Ebola-related tweets and web searchers. The study looked at the time period of September 29th 2014 to October 31st, 2014 and sought to examine Internet searchers and tweets. The authors were able to obtain TV news coverage data by monitoring the number of Ebola-related news videos that had appeared on two main news networks in the US. They found that there was evidence to suggest that TV news coverage had led to an increase in Ebola-related tweets as well as Internet searches. This is because there was a correlation between TV news coverage and an increase in Ebola related tweets.

Rodriguez-Morales, Castañeda-Hernández, and McGregor (2014) in a correspondence for the journal Travel Medicine and Infectious Disease, noted that social media tools such as Twitter have greatly increased the communication potential between networks of people who were previously disconnected. In the study by Towers *et al.* (2015) described above, media coverage related to Ebola was analysed, whereas Rodriguez-Morales, Castañeda-Hernández, and McGregor (2014) examined social media activity using Topsy, which was a travel-related infection at the time. They noted that there had been an 11-fold increase in Ebola-related tweets with a peak of 1.5 million tweets on the 30th September 2015, which coincided with diagnosis of Ebola for an individual who travelled to the United States from Liberia at that time (McCarthy, 2014).

Stefanidis *et al.* (2017) examined how Zika, which has recently been a major international issue, was communicated on Twitter. The Zika virus is a mosquito-borne flavivirus (Aedes) that shows a broad range of potential symptoms ranging from arthralgia, mild fever, and headaches, among others (Ioos, 2014). The study retrieved a total of 6,249,626 tweets across a 12-week period from December 2015 to March 2016. The study examined the location of the users that were tweeting as well as the types of users, and found that some of the most influential accounts were related to the Centers for Disease Control and Prevention (CDC) and the World Health Organization (WHO). The data were analysed using network tools that support spatiotemporal analysis. Moreover, the study found that the interest of Twitter users in relation to Zika reflected how the virus was spreading across the world. The authors also found that tweets began to mention pregnancy and abortion as people became more aware of Zika and its effect on pregnancy, and as more public figures would comment on the disease.

A brief report published in the American Journal of Infection Control by Fu *et al.* (2016), computationally analysed 65,547 English tweets on Zika over a 3-day period using the automated method of topic modelling. They found that tweets fell under 5 key themes: the impact the outbreak was having on society, responses to Zika such as from the general public, government and private sector, pregnancy and microcephaly, the transmission of Zika, and a theme with case reports of Zika. A further brief report examining Zika by (Dredze, Broniatowski, & Hilyard, 2016) found that a number of pseudo-scientific claims were shared on Twitter from January 1st to April 29th 2016.

2.9.1 Summary

In summary, this section provided an overview of a number of studies that were conducted on Twitter related specifically to infectious disease outbreaks such as swine flu, Ebola, and the Zika virus. It appeared that many of these studies would utilise a qualitative approach to analysing tweets. The studies that analysed infectious disease outbreaks are brought together in a comparator table, 2-1 below. The key findings, methods, strengths, and limitations of the studies are highlighted.

Table 2-1 Summary of studies analysing Twitter data related to infectious disease outbreaks

Study	Topic	Key findings	Methods	Strengths	Limitations
Chew and Eysenbach (2010)	Swine flu	Twitter has the potential to be used to estimate disease outbreak patterns, and can be used to gain real-time insights into public concerns. Of the tweets analysed, 6 main themes were found to emerge which consisted of: resource, personal experience, personal opinion and internet, Jokes/Parody, Marketing, and spam.	Mixed Methods: analysis (qualitative) and machine learning for automated coding and sentiment analysis (quantitative).	One of the earliest studies to demonstrate the potential of Twitter data for disease surveillance and public views and opinions.	It was not possible to define the demographics of the sample of tweets that were analysed. Tweets that were retrieved could have been tweeted from anywhere in the world. So it was not possible to examine views from specific geographical areas.
Signorini, Segre and Polgreen (2011)	Swine flu	Twitter can be used to estimate the incidence of disease activity, and to better understand the sentiment of the general public.	A quantitative approach was utilised and an estimation model was built.	Further validated the use of Twitter for estimating disease activity, and gaining real-time insights.	Noted that it was not possible to validate the results public views on Twitter, i.e. by comparing them to offline data such as surveys.
Kostkova, Szomszor, and St. Louis (2014)	Swine flu	Twitter has the potential to predict the peaks in outbreak before traditional methods. Reputable sources were more successful in communicating on Twitter when WHO declared swine	A quantitative approach was utilised which involved text classification and time series analysis.	A good sample of tweets were originally obtained (3 million tweets).	Twitter is not representative to the offline population, and there may be certain biases in reporting.

		flu in 2009 to be a pandemic.			
Oluwafemi, Elia, and Rolf (2014)	Ebola	Most tweets and retweets on Ebola contained medical misinformation.	A qualitative approach was applied because tweets were manually grouped together.	Among one of the first empirical studies to indicate that there may have been medical misinformation present on Twitter.	Lacks rigour in reporting methodological considerations. There is no indication of the number of tweets that were analysed and how they were captured.
Jin ***et al.*** **(2014)**	Ebola	Found that during the 2014 Ebola epidemic that there were rumours and conspiracy theories on the platform.	A computational approach was utilised to develop a model in order to identify and study rumours.	Further added to the body of knowledge by highlighting specific fears that were circulating on the platform.	The identification and reporting of the results lacked interpretation. Some of the rumours could be classed as humour.
Odlum and Yoon (2015)	Ebola	It was found that Twitter was a useful platform to inform public health. The study found that Twitter had the potential for outbreak surveillance.	A quantitative approach was applied because time series analysis was used, and tweets were classified using natural language processing.	Identified a number of public health concerns, and was able to examine their geographical location.	There are limitations in the generalisations that can be made because a single language was examined (English).
Scheinfeld, Bernhardt, Wilcox, and Suran, (2015)	Ebola	Provided new insight into the types of information the public required during the outbreak.	A quantitative approach was utilised because machine learning was utilised to analyse tweets.	Demonstrated how Twitter chats could be analysed to gain insight into public views and opinions.	The study was limited because the findings were based on a single Twitter chat which took place in English.
Kalyanam, Velupillai, Doan, Conway, and Lanckriet (2015)	Ebola	Provided new insights into the information propagation of false and credible	A quantitative approach using machine-learning classifier was utilised.	Study outlined a methodology of assigning tweets	Only tweets which contained hashtags were captured. When the classifier was trained,

		information on Twitter.		into two sets: credible and false.	tweets were manually labelled, and there was no verification to see whether labels were assigned appropriately.
Towers *et al.* **(2015)**	Ebola	Correlation between TV-coverage on Ebola and an increase in Ebola-related tweets.	A quantitative method was utilised.	First empirical to find a positive correlation between TV-coverage on Ebola and an increase tweets.	Only focused on American TV coverage, and the results are not applicable to other countries.
Rodriguez-Morales, Castañeda-Hernández, and McGregor (2014)	Ebola	Highlighted there was a large peak in tweets when a patient with Ebola was brought to US during the 2014 epidemic.	Utilised a quantitative method to identify peaks in tweets.	Highlighted how a major news story was able to lead to an increase in tweets.	Performed no content analysis on the tweets that were identified to have peaked.

2.9.2 Gaps in current knowledge

In regards to some key trends across the studies, as shown in Table 2-1, studies on infectious disease outbreaks on Twitter utilise a wide range of different methods, and many of these studies are primarily quantitative based. An interesting gap in knowledge is the lack of empirical studies which attempt to apply an in-depth qualitative methodology for the analysis of tweets on infectious disease outbreaks.

Research conducted on other health topics have made use of in-depth qualitative methods (Hewis, 2015; Shepherd, Sanders, Doyle, and Shaw, 2015), but these types of studies (i.e. those utilising in-depth methods) were not found to have been conducted for either the swine flu pandemic and/or the Ebola epidemic. Additionally, previous research on Zika appears to have been conducted utilising quantitative methods. Moreover, there is also a gap in knowledge on whether Twitter data is a useful platform for gaining in-depth qualitative insights into emerging disease outbreaks in comparison to traditional qualitative methodologies.

Although, a number of studies have noted the potential of Twitter data to provide insights into public views and opinions and/or validated its usefulness for disease surveillance, no previous empirical study assessed the utility of Twitter for gaining in-depth qualitative insights. Moreover, although previous research has discussed the strengths and weaknesses of using Twitter data for gaining insight into public views and opinions and for disease surveillance, no study has specifically considered the utility of Twitter data for gaining in-depth qualitative insights.

2.10 Crisis Communication using Social Media Data

The previous section outlined how Twitter has been used to monitor general health and identified a number of gaps in current knowledge. This section will report on studies that have investigated Twitter for health crisis communication. This is important to include in this review of existing literature because Twitter has emerged, as a fruitful platform for research into crisis communication and the methodological considerations of previous work in this area is important to review.

Studies based on health crisis and risk communication are centred on disease surveillance and early warning systems using Twitter data. Using data from social media platforms, such as Twitter, to build early warning systems for infectious disease outbreaks and natural disasters has gained interest in recent years (Chew and Eysenbach, 2010). The research papers below consist of literature reviews as well as original empirical research.

Salathé and Khandelwal (2011) examined Twitter to see whether it would be possible to detect sentiments on vaccinations across the US population in order to identify target areas for increasing vaccination intervention. Traditional methods which measure public sentiments on vaccinations and plot the distribution of these sentiments across the population are very difficult to implement and may consume many resources. However, it is important to understand sentiments on vaccinations as they can greatly influence an individual's decision. Salathé and Khandelwal's (2011) approach was validated by the identification of a strong correlation between Twitter sentiments and the CDC estimates on vaccination rates based on telephone surveying. The authors also found that the flow of information is most prevalent between users who express the same sentiments, and this is reduced if users have different views. They concluded that social media such as Twitter can provide vast amounts of data and are a cost-effective approach to identifying and targeting areas for vaccination intervention compared to more traditional methods. The limitations of using Twitter are that its users may not be a representative sample of the general population, different messages may be interpreted differently by different people, and sentiment analysis may misclassify tweets, for example, if a tweet expressed humour or sarcasm, this may get classified as being negative.

Bernardo, Rajic, Young, Robiadek, Pham, and Funk (2013) conducted a review of the literature examining how web search queries had been used for disease surveillance. The study was undertaken in order to inform future work on early detection of foodborne illnesses. In total, the study identified 32 research articles as well as 19 reviews. The study found that many of the papers (n = 21) noted that social media surveillance had similar performance to traditional surveillance programmes. The authors noted that one of the strengths of social media surveillance programs was related to 'rapid detection' and effectiveness. However, a weakness that the study identified related to the potential of locating false positives as well as false negatives. For example, a person may report a case of Ebola incorrectly, either unintentionally or for malicious purposes, and this would be a false positive. There may also be instances where a person does have Ebola, but may reside in a location where Ebola is not prevalent and so an algorithm may decide that the person does not have Ebola, and this would be a false negative. Moreover, the study stressed that social media surveillance programs should be used in association with more traditional methods of surveillance.

Yom-Tov, Borsa, Cox, and McKendry (2014) employed custom-built algorithms that raise the alert when possible outbreaks of communicable diseases were suggested via Twitter and search engine queries. Mass gatherings were also studied, including nine musical festivals and one religious event (i.e. the Hajj in Saudi Arabia) in 2012 for 30 days. Mass gatherings are a potential source of infectious disease spreading, because of the close proximity of people to each other. All Twitter postings and queries made to the Bing search engine were extracted. However, no major infectious disease outbreak was identified, although there was agreement by users reporting a cough during one of the festivals. However, traditional surveillance programs are challenged, as people gather and disperse quickly. Surveillance through social media and the internet during mass gatherings can act as a rapid and more effective surveillance program for infectious diseases. This is due to the challenges of incubation period and dispersion of participants. A weakness of using Twitter for disease surveillance compared to traditional methods is the chance for false positives.

In a similar study, Rutsaer *et al.* (2014) examined new media and web technologies and whether these could improve food risk and benefit communication. The study considered the views of 38 stakeholders and 33 experts whose interest was in the field of food safety. In-depth interviews were conducted across six countries. Strengths, weaknesses, opportunities, and threats (i.e. a SWOT analysis) of food risk and benefit communication were also examined. The findings of the research indicated that both stakeholders and experts saw a future in social media for food risk and benefit communication. The strengths of social media were reported as coming from their ease of access and speed, and the weaknesses arising from the chance of false positives, no filtering, the lack of trust, and information overload. It was also noted that social media may create unneeded hysteria and mass panic around food crises, mainly due to the emotional behaviour of users and lack of filters. However, the authors noted that “With approximately two billion people having access to the internet in 2012… health policy makers are strongly recommended their use alongside traditional outreach models” (p.93).

These reviews and original research papers have shown that the main limitation of using Twitter for early warning systems and detection is that of false positives and or false negatives, and that the main advantage is faster detection. Researchers have recommended that early warning systems be used alongside traditional methods rather than replace them outright because of the possibility of false positives. There are also studies that have used social media data for non-health-related crisis communication and decision making such as riots, earthquakes, storms, emergency situations and crisis events, which will be outlined next.

Procter, Vis, and Voss (2013) conducted a study on an experiment which utilised a computationally-assisted methodology to analyse a large dataset of tweets sent during the 2011 riots in England. They noted that, although Twitter could be used to spread rumours, it could also be used for the correction of such rumours. Citing Mendoza, Poblete, and Castillo, (2010), they noted that users tend to deal with rumours that are true differently than rumours that are false. True rumours are confirmed 90% of the time and false rumours are either questioned or denied 50% of the time. More generally, the authors noted that the use of Twitter for social science research poses a number of methodological challenges. Firstly, they noted the possibility of sampling bias (i.e. that not all tweets related to the riots would be retrieved), which could distort the findings. This is because it is very likely that users talked about the riots using different search queries to the ones used to generate the dataset. Another methodological limitation was in the ability to draw inferences from the results (i.e. whether Twitter was used to incite unlawful acts), as their sampling frame excluded Direct Messages (DMs). In other words, the authors questioned whether analysing tweets that were posted publically could alone be used to draw conclusions when it is likely that members of the public exchanged information via Direct Messages (DMs) (Purohit *et al.*, 2013).

Mendoza, Poblete, and Castillo, (2010) looked at how Twitter users were behaving during an earthquake in Chile in 2010 to understand better how people react during emergency situations. The study analysed tweets before and after the disaster, and specifically examined how users were sharing false as well as confirmed news. The aim of the study was to assess whether Twitter was a reliable information source during extreme circumstances. The study noted that there was a difference between how users would interact with real news as opposed to false news, that is, it was found that when users shared tweets containing rumours, these would be questioned by users.

Ghahremanlou, Sherchan, and Thom (2015) noted that a number of algorithms, such as OzCT, would automatically identify the locations of Twitter users based on tweet content. It would therefore be possible to conduct surveillance without the need for tweets to have had geolocation information. They noted that, during a disaster, situation microblogging tools such as Twitter could act as important communication platforms for the dissemination of information.

Gupta, Joshi, and Kumaraguru, (2012) conducted an analysis on three main events that occurred in 2011: an earthquake in Virginia, riots in England, and Hurricane Irene. They noted that Twitter is a key social media platform used to share opinions and information. They used a network measure of degree centrality in order to identify influential users. Degree centrality looks at the number of nodes connected to a node, for example, on Twitter this would be the number of Twitter users that follow a user, and a higher number of nodes would indicate a higher influence score (Opsahl, Agneessens, Skvoretz, 2010). They were able to demonstrate that influential Twitter users represented all of the topics and opinions that were shared within the community with 81% accuracy on average. The study concluded that the most central Twitter users tend to represent content that is shared by the entire Twitter community. They noted that in order to understand a particular community, it would be sufficient to monitor and analyse the users with high influence scores in a community rather than all users.

Researchers have also studied Twitter data related to major weather events. For example, Lachlan, Spence, Lin, Najarian and Greco (2015) examined the types of content users were sharing during a major weather event (a storm named Nemo which affected North America in February 2013). They found that emergency services had not utilised Twitter to its full potential, as useful information tends to be easier to find when using hashtags that have keywords that are local. However, they did noted that messages are difficult to locate during a weather emergency, as there may be many tweets that are sent using the hashtag and keywords from outside the affected areas. They recommended that official responders, as well as emergency service agencies, should utilise hashtags that may emerge organically (i.e. from Twitter users themselves, as an emerging event unfolds, so that they can locate information which can be used by the public).

Social media platforms have altered the way in which people communicate with each other (Cameron, Power, Robinson, and Yin, 2012). This is because, when an emergency situation arises, information can be available from members of the public and be used as intelligence by emergency services and to allocate resources. These new sources of information will not entirely replace existing ones. However, they will be able to provide new data sources that could have a number of applications during an emergency. Cameron, Power, Robinson, and Yin (2012) noted that social media has the potential to play a role before, during, and also after, a crisis event.

Simon, Goldberg, Aharonson-Daniel, Leykin and Adini (2014) looked at Twitter activity related to a crisis event corresponding to an attack on the Westgate mall in Kenya in 2013. The concluded that Twitter had become a key communication channel between a number of stakeholders such as government, emergency responders, and the general public as the crisis was unfolding. The study examined 67,849 tweets and four categories of hashtags. During the crisis, emergency personnel were observed utilising social media networks for communicating information with the general public as well as among themselves.

This subsection has highlighted how Twitter data has been utilised for crisis situations and natural disasters such as earthquakes, riots, disease outbreaks, and major weather events. The next section provides insight into studies which use in-depth methods to analyse tweets.

2.11 Studies using In-Depth Qualitative Methods to Analyse Tweets

By bringing together literature related to Twitter and health it was found that the majority of studies outlined in this literature review utilised quantitative methodologies to analyse Twitter data (Salathé and Khandelwal, 2011; Signorini, Segre and Polgreen, 2011; Bosley, Zhao, Hill, and Shofer, 2012; Kostkova, Szomszor, and St. Louis, 2014; Santos and Matos, 2014; Jin *et al.*, 2014; Young, Rivers and Lewis, 2014; Cavazos-Rehg *et al.*, 2015; Odlum and Yoon, 2015; Liu, Mei, Hanauer, Zheng, and Lee, 2016; Kendra, Karki, Eickholt, and Gandy, 2015; Stefanidis *et al.*, 2017). Few studies have utilised a purely qualitative approach (Scanfeld, Scanfeld and Larson, 2010; Robilard *et al.*, 2013; Oluwafemi, Elia, and Rolf, 2014). This literature review covers articles which may have analysed Twitter data on any topic, including non-health related topics, using in-depth methods such as framework analysis and/or thematic analysis.

A study which used framework analysis as a variant of thematic analysis (Srivastava and Thomson, 2009) looked at Twitter activity during the Vancouver riots that took place on June 15th 2011 (Burch, Frederick, Pegoraro, 2015). Tweets were retrieved based on whether they contained specific keywords or hashtags. The total dataset consisted of 1,226 tweets with 1,026 unique users. By performing a textual thematic analysis of tweets, the authors found that the majority of material related to five themes: 'fandom', 'riot propagation', 'global perspectives', 'shame on Vancouver', and 'real fans *vs.* idiots'. Burch, Frederick, and Pegoraro (2015) concluded that the identification of these themes highlight Twitter as means of news and information and as an channel for influencing public opinion. A distinct feature of this study was the number of quotations that were included. It appears that the tweets within the study were not anonymised and the users could be identified. However, as the tweets related to a sporting event, the authors may have deemed the risk to the individuals as low, although this still raises importance ethical concerns, which will be addressed in Chapter 3.

Hewis (2015) noted that Twitter may be able to provide unique insights into the magnetic resonance imaging (MRI) patient experience, and utilised a qualitative methodology to analyse tweets. During one calendar month, 6,471 tweets were extracted and 464 tweets were categorised. The search strategy was that all tweets containing the phrase 'MRI', or 'magnetic resonance imaging' between the 1st and the 31st May 2015 were retrieved, and this search was conducted using the native Twitter search engine. Hewis (2015) found that the majority of material related to three themes: MRI appointment, scan experience, and diagnosis. Hewis (2015) noted that the study demonstrated that patients suffering from MRI will tweet about their experience and Twitter is a viable platform to conduct research into patient's experience within the medical radiation sciences. They noted that, due to the short length of tweets, there were challenges with depth and the potential of incorrectly interpreted tweets.

Shepherd, Sanders, Doyle, and Shaw (2015) noted that there is a lack of empirical research examining the potential role of social media by individuals with mental health problems. They set out to examine a single discussion in order to assess the role that Twitter can play as a medium for interpersonal communication. They searched Twitter for potential mentions of the keyword '#dearmentalhealthprofessionals'. Duplicate tweets were removed, and the remaining tweets (515) were reviewed and analysed using thematic analysis. Four overarching thematic headings emerged (Shepherd, Sanders, Doyle, and Shaw, 2015):

- "The impact of diagnosis on personal identity and as a facilitator for accessing care" (p.4)
- "Balance of power between professional and service user" (p.4)
- "Therapeutic relationship and developing professional communication" (p.4)
- "Support provision through medication, crisis planning, service provision and the wider society"(p.4)

The study demonstrated the utility of Twitter as a platform where individuals with an experience of mental disorder can share information, develop understanding, and offer feedback to mental health service providers. A useful feature about the qualitative approach was that the results were rich in detail, because a number of tweets were provided as examples under the thematic headings.

Heaivilin, Gerbert, Page, and Gibbs (2011) investigated whether it was possible to identify tweets on dental pain. Twitter was searched using a number of queries related to toothache, and from this a sample of a thousand tweets were extracted. In the next step, duplicate tweets from individual users were removed, and a total of 772 tweets remained. The authors then performed thematic analysis on tweets and coded up to 300 tweets. At this stage, they noted that thematic saturation was likely i.e. that it was unlikely new themes would emerge from the analysis of data (Heaivilin, Gerbert, Page, and Gibb, 2011). Nevertheless, a diverse range of themes and sub-themes were found to emerge related to dental pain and thus it was concluded that Twitter could act as a new instrument for those within the health sector to communicate with patients.

In the next section, the studies outlined in the literature review will be brought together in a summary table. This will allow for the direct comparison of the research methods used by each of the studies.

2.12 Summary Table

In order to compare the methods used by each of the studies and to ascertain the number of tweets that were captured and analysed, the studies using Twitter data as a primary data source for health-related-research are brought together in a summary table (Table 2-2). The sub-selection criteria for generating this table were that the study presented original results and used Twitter as a primary data source. The primary selection criteria for locating literature were as simple as “Twitter” alongside a health condition; for example, “bird flu” AND “Twitter”, or “HIV” AND “Twitter”, and as general as “Twitter” AND “health”. For crisis and risk communication literature, the search terms were as simple as “Twitter crisis”, “Twitter” AND “risk”, “Twitter” AND “decision making”. Additionally, keywords such as “social media” AND “swine flu”, “social media” AND “ebola”, and “social media” AND “healthcare” were used. The majority of studies in Table 2-1 (with the exception of Signorini, Segre and Polgreen, 2011 and Jin *et al.*, 2014) provided the number of tweets that were examined by each study and which ranged from 2,000 to 2,000,000.

Table 2-2 Summary of health research literature related to Twitter (comparative table

Study	Topic	Overall method	Methodology/ No. of tweets examined (if applicable)	Finding
Chew and Eysenbach (2010)	Swine flu pandemic during 2009. The aim was to see whether Twitter could be used to extract public views and opinions on a larger scale.	Mixed Method	Content analysis and machine learning to analyse data. The study retrieved over 2 million tweets from 1st May to the 31st December. The keywords that were used to retrieve the data consisted of swine flu, swineflu, and/or H1N1.	Tweets which were based on H1N1 were related to sharing trustworthy information as well the views and opinions of users. Twitter was a valuable source of information as the pandemic was unfolding.
Kostkova, Szomszor, and St. Louis (2014)	Swine flu outbreak from 2009 in order to highlight the usefulness of early warning systems.	Quantitative	Methods included text-classification and time series analysis. A total of 3 million tweets were retrieved from May 7th 2009 to December 2009.	Twitter was able to sample large populations and had the potential to predict infectious disease outbreaks.
Signorini, Segre and Polgreen (2011)	Tweet content of the H1N1/09 (swine flu) outbreak to see whether it was possible to track the spread of the virus.	Quantitative	Building of an estimation model, in which a large number of tweets were gathered which matched a set of predefined keywords. The exact number of tweets used in the study not provided.	Twitter had potential to be used for following the outbreak to understand public views, but that it could also be utilised by having the potential to be used for tracking and detecting the spread of the virus faster than traditional methods.
Santos and Matos (2014)	Quantitative methods to better understand whether content generated by Twitter could be used to develop a model that would provide instant feedback	Quantitative	The method of analysis for this study was to create an estimation model. This study was based on a dataset of 2,704 tweets which were sent from Portugal.	When tweets and search engine queries are combined that this combination has the potential to predict as well as produce an estimate of influenza-related illness in Portugal. A key advantage of using social media and web queries

	on the levels of incidence of flu in Portugal.			was that they tend to be faster in comparison to traditionally published reports.
Jin *et al.* (2014)	To test the hypothesis that rumours as well as misinformation were present on Twitter during the 2014 Ebola epidemic.	Quantitative	Computational methods used to create a model to classify a number of tweets that were gathered based on a large number of keywords and hashtags. Exact number of tweets that were retrieved were not provided by the authors.	In addition to a number of news stories that were shared there were also a number of conspiracy theories as well as rumours that were found to have been shared on social media.
Oluwafemi, Elia, and Rolf (2014)	Twitter was utilised in order to retrieve tweets that were in English and mentioned a number of keywords related to Ebola.	Qualitative	Qualitative approach to manually group tweets. Utilised Twitter's advanced search in order to seek tweets that were related to Ebola and had been sent from Guinea, Liberia, and Nigeria. The tweets were sent from the 1st to 7th September 2014.	Rumours were shared on Twitter related to curing Ebola. This false information was only corrected occasionally by the Nigerian government. A recommendation derived from the research was that governments should utilise Twitter in order to disseminate accurate information.
Scanfeld, Scanfeld and Larson (2010)	Examined tweets which mentioned the keyword antibiotics. Aim of the study was to discover whether there were potential misunderstandings and/or misuses of antibiotics.	Qualitative	The method used to analyse data was content analysis, which is a qualitative research methodology. In this study 1,000 tweets were retrieved which mentioned the keyword antibiotics.	Twitter was a valuable tool for health authorities, and could be used to gather Twitter data in real-time. It was important for health authorities to have a basic understanding of social media websites such as Twitter.

Bosley, Zhao, Hill, and Shofer (2012)	Examined how users would share information on Twitter related to cardiac arrest.	Quantitative	The tweets were analysed using computational methods such as text-classification. This study used a number of keywords which relate to cardiac arrest to retrieve 62,163 tweets.	It was possible to retrieve tweets and to filter the tweets in order to identify how users were sharing information related to cardiac arrest.
Robilard *et al.* (2013)	Examined how age-related disease information, for example, dementia is shared on Twitter.	Qualitative	The method of analysis related to utilising machine learning to classify tweets. The study examined Twitter for a 24-hour period and a total of 9,200 tweets were retrieved.	A large amount of tweets were related to sharing news rather than personal information, and the users who were sharing information were health professionals. The content that was shared was related to studies that had been published recently on predicting the incidence of dementia.
Young, Rivers and Lewis (2014)	Examined whether Twitter users would share information related to sexual risk behaviour on Twitter.	Quantitative	Computational methods were utilised by performing geographical analyses. Total of 553,186,062 tweets were gathered and then filtered in order to discover whether the tweets would mention HIV-related behaviours.	A significant positive correlation between HIV Twitter activity and the amount of HIV cases in the United States. The authors noted that social media had the potential to map the incidence of HIV.
Salathé and Khandelwal (2011)	Examined whether it was possible to identify sentiments related to vaccinations in the US population. This would allow for the targeting of specific geographical areas in order to increase vaccination intervention.	Quantitative	Computational analysis methods were utilised which included sentiment analysis, and correlation analysis. The study captured 477,768 tweets and it was found that of these tweets 318,379 were related to the A (H1N1) vaccine.	Social media websites such as Twitter have the potential to provide a large amount of data which can be feasibly obtained in order for vaccination intervention.
Cavazos-Rehg *et al.* (2015)	The study sought to examine the themes and	Quantitative	Utilised methods to analyse the sentiment and themes of the results. A	High frequency of tweets which were related to marijuana were

	sentiment that would emerge from tweets related to marijuana.		large number of tweets were retrieved (14.5 billion), which matched a number of predefined keywords. Of these 7,000 were examined in the study.	positive in sentiment. Moreover, they noted that this information could be utilised in order to target users at risk of marijuana use.
Odlum and Yoon (2015)	This study aimed to demonstrate the potential of Twitter for disease surveillance.	Quantitative	The methods of analysis in the study were computational which included natural language processing. A total of 42,236 tweets were retrieved.	It was possible to utilise Twitter data in order to inform public health education.
Liu, Mei, Hanauer, Zheng, and Lee (2016)	This study aimed to examine and monitor a number of keywords related to diabetes.	Quantitative	A number of keywords were used in order to retrieve 29.6 billion tweets and from this a total of 1,368,575 tweets were extracted.	Twitter has potential for gaining insight into how members of the public converse about diabetes.
Kendra, Karki, Eickholt, and Gandy (2015)	This study examined Twitter data to see if tweets related to health topics could be extracted as well as to see if users were conversing about antibiotics.	Quantitative	Utilised computational analysis methods which included a network classifier. A total of 591,091 tweets from were retrieved.	There were a many users that would tweet about antibiotics. Moreover it was also found that users would discuss antibiotic resistance the most.
Stefanidis *et al*. (2017)	This study sought to examine Zika discussions on Twitter.	Quantitative	Utilised computational methods such as network analysis and geographical analysis. A total of 6,249,626 tweets were retrieved from a 12-week time frame.	Most influential Twitter accounts were that of the Centres for Disease Control (CDC) and the World Health Organisation (WHO).
Shepherd, Sanders, Doyle, and Shaw, (2015)	This study sought to examine the potential role of social media platforms by individuals who suffer from mental health issues.	Qualitative	Made use of thematic analysis to analyse data. A total of 515 tweets were retrieved and analysed.	Four overarching themes emerged from the analysis of data.
Hewis (2015)	The aim of this study was to examine the potential of gaining insight on	Qualitative	Study utilised thematic analysis. A total of 6,471 tweets were retrieved and, of these, 464 tweets were analysed.	MRI patients did tweet about their experiences and that Twitter cold be used in order to conduct research.

	patient MRI experiences using Twitter data.			
Heaivilin, Gerbert, Page, and Gibbs (2011)	Study examined dental pain discussions on Twitter	Qualitative	Study made use of thematic analysis. A total of 300 tweets were coded.	Twitter users share information related to dental pain on Twitter.

As shown in table 2-2 above, the types of health related topics that were captured and analysed on Twitter varied from marijuana use (Cavazos-Rehg *et al.*, 2015), sentiments on vaccinations (US Cavazos-Rehg *et al.,* 2015), sexual risk behaviours (Young, Rivers and Lewis, 2014), age-related disease information (Robilard *et al.*, 2013), cardiac arrest (Bosley, Zhao, Hill, and Shofer, 2012), dementia (Robilard *et al.*, 2013), antibiotics (Scanfeld, Scanfeld and Larson, 2010), Ebola (Jin *et al.*, 2014; Oluwafemi, Elia, and Rolf, 2014; Odlum and Yoon, 2015), and swine flu (Chew and Eysenbach, 2010; Signorini, Segre and Polgreen, 2011; Kostkova, Szomszor, and St. Louis, 2014). The majority of studies used a quantitative approach (Salathé and Khandelwal, 2011; Signorini, Segre and Polgreen, 2011; Bosley, Zhao, Hill, and Shofer, 2012; Kostkova, Szomszor, and St. Louis, 2014; Santos and Matos, 2014; Jin *et al.*, 2014; Young, Rivers and Lewis, 2014; Cavazos-Rehg *et al.*, 2015; Odlum and Yoon, 2015). Tweets were, in most cases, placed into categories and then categorised computationally by utilising machine learning which involves assigning labels to data points, and then allowing a classifier to automatically code a dataset (Pennacchiotti, and Popescu, 2011).

Other studies selected a purely qualitative methodology, in which tweets were categorised manually (Robilard *et al.*, 2013; Scanfeld, Scanfeld and Larson, 2010; Oluwafemi, Elia, and Rolf, 2014; Shepherd, Sanders, Doyle, and Shaw, 2015; Hewis, 2015). One study (Chew and Eysenbach, 2010) employed a mixed methods approach whereby manual content analysis was performed on tweets alongside automated content analysis, and sentiment analysis. Many of the studies utilised quantitative computational methods for analysis, such as automated text classification, machine learning, and natural language processing (Bosley, Zhao, Hill, and Shofer, 2012; Robilard *et al.*, 2013; Jin *et al.*, 2014; Kostkova, Szomszor, and St. Louis, 2014; Odlum and Yoon, 2015). Other studies noted that they involved some form of sentiment analysis (Salathé and Khandelwal, 2011; Cavazos-Rehg *et al.*, 2015). Others studies would utilise computational quantitative methods such as building estimation and prediction models (Signorini, Segre and Polgreen, 2011; Santos and Matos, 2014), qualitative methods such as content analysis (Scanfeld, Scanfeld and Larson, 2010; Oluwafemi, Elia, and Rolf, 2014) and thematic analysis (Hewis, 2015; Sanders, Doyle, and Shaw, 2015).

The findings of this literature review are similar to a systematic review published in 2017, after the original version of this literature review was written in 2015, which sourced and synthesised peer reviewed articles analysing Twitter data for health research (Sinnenberg, 2017). Of the 137 articles which met the eligibility criteria, they found that previous studies utilised methods such as content analysis (56%), surveillance (26%), engagement (14%), recruitment (7%), intervention (7%), and network analysis (4%) (Sinnenberg, 2017). They found that at least 20% of previous studies utilising Twitter data for health were based on analysing infectious disease outbreaks. However, the study did not comment on the effectiveness of Twitter data for in-depth qualitative insights and/or highlight a lack of in-depth methods being applied to Twitter data. Moreover, the study did not distinguish between studies using content analysis supported by machine learning, and those which were purely qualitative.

2.13 Synthesis of Literature

The primary aim of this chapter is to review relevant literature, an aim that follows on from this was to examine the methods employed in previous research on Twitter, particularly in relation to health. The literature review found that there are a variety of methods that have been utilised to gather and analyse Twitter data. However, certain aspects were similar across the studies, as many of the studies gathered Twitter data via the public API ecosystem, i.e. the Streaming or Search APIs, as tweets were retrieved, due to feasibility. The time periods selected to obtain data varied, from 24-hour intervals to several months. The commonalities between studies that utilised Twitter data for crisis and risk communication for Twitter were speed and accessibility and these formed the main motives for using social media for these purposes. The barriers faced were lack of trustworthiness, lack of filtering, and the possibility for false positives. The studies on crisis and risk communication suggested that Twitter may be a good platform for early warning and detection of infectious disease outbreaks. However, using Twitter data will face the same challenges that search query detection systems, such as Google flu trends, have faced, and correlation does not always imply causation (Lazer, Kennedy, King, and Vespignani, 2014): there could be a number of reasons why members of the public may search for information related to flu, and a large occurrence of tweets and/or search queries may not necessarily imply that that there is an increase in fthe disease. These systems will also run the risk of over-predicting occurrences of diseases such as flu. Google flu trends, for instance, have over-predicted the occurrence of flu (Arthur, 2014), and did not disclose their methodology, specifically their algorithm, which contained the search queries that Google uses. These systems may rely on the fallacy that correlation implies causation, and thus it may only be possible to know if they are defective *after* there has been a misinterpretation. In the case of using Twitter data, this could occur when large numbers of tweets from a specific geographical location arise out of general interest rather than the outbreak of a disease. Therefore, early warning systems should not be relied upon exclusively, and should be employed alongside more traditional methods, as suggested by most of the authors.

Another key aim of this literature review was to identify other emergencies or crisis instances in which Twitter data were analysed. It was found that research has analysed tweets on riots (Procter, Vis, and Voss, 2013), natural disasters (Mendoza, Poblete, and Castillo, 2010; Lachlan, Spence, Lin, Najarian and Greco, 2015), and crisis events (Gupta, Joshi, and Kumaraguru, 2012; Simon, Goldberg, Aharonson-Daniel, Leykin, and Adini, 2014).

The studies related to natural disasters reported that Twitter was an important means of communication across government, emergency services, and the general public during a crisis (Cameron, Power, Robinson, and Yin, 2012; Simon, Goldberg, Aharonson-Daniel, Leykin, and Adini, 2014). The studies also noted that social media has changed the way people communicate (Cameron, Power, Robinson, and Yin, 2012). Other studies noted that emergency management agencies in certain situations have underutilised Twitter (Lachlan, Spence, Lin, Najarian and Greco, 2015). Relatively few studies utilised thematic analysis as a method of analysing tweets (Burch, Frederick, and Pegoraro, 2015; Hewis, 2015; Shepherd, Sanders, Doyle, and Shaw, 2015). The studies that utilised thematic analysis reported results by listing themes to emerge from the data and by providing extracts and quotes from the tweets to be included alongside the themes.

Difficulties arise in replicating some of the research outlined above, as Twitter does not permit data sharing. However, through methodological transparency, it is possible to know how the data were created (Vis, 2013), i.e. by understanding what data were gathered during which time periods, and which metadata were or were not selected. This would allow researchers to understand the findings better and replicate the results using a new set of Twitter data. However, not all studies described above contained precise details of their methodology.

2.14 Limitations of Existing Research

The majority of studies that examined Twitter data in relation to health and crisis communication lacked a theoretical basis and may have neglected the opportunity to relate theoretical concepts to the results. This may have led to more richness within the results. Existing studies have tended to utilise a quantitative approach and few have utilised thematic analysis and/or produced qualitative reports. This could mean that the interpretations drawn will be limited in scope, because simply identifying and/or classifying tweets into predefined categories will provide fewer insights into the events taking place in the world and on Twitter. For example, in the context of the study by Jin *et al.* (2014), which looked at conspiracy theories in the context of Ebola, if the study had incorporated a more qualitative approach, there could have been more critical insight into whether what was being tweeted was a rumour, whether users were being humorous, and/or whether there was a feature of the online world which made users engage with this content. It appeared that only a limited number of health-based studies would provide qualitative insights into tweets. Traditional qualitative reports, such as outlining themes from tweets and extracting quotes and interpreting the discussions that were taking place, are missing from previous studies. These methods are important and helpful because they provide context to the tweets, and are richer overall, because they will often contain the tweets that are associated with each of the themes. Although Twitter data may be large in size, reducing a dataset would make it possible to extract qualitative insights. Therefore, studies that do not provide qualitative insights into social media data may be lacking a certain richness, and may fail to highlight the context in which tweets occur in. It also appears that previous studies have retrieved data via the free APIs provided by Twitter, which are a sub-sample of the entire Twitter stream. Therefore, many of the previous studies have analysed a sample from a sub-sample (Chew and Eysenbach, 2010; Signorini, Segre and Polgreen, 2011; Jin *et al.*, 2014; Oluwafemi, Elia, and Rolf, 2014; Odlum and Yoon, 2015), which has the potential to render the datasets incomplete.

2.15 Lack of Studies and Knowledge Gaps

The studies in Table 2-2 have validated Twitter as a valuable source of information for health-related topics and global crisis events. However, they also highlighted a lack of empirical research that has examined the swine flu pandemic and the Ebola epidemic using an in-depth qualitative methodology. There is also a lack of studies that utilise thematic analysis to analyse infectious disease outbreaks on Twitter, although thematic analysis was applied to tweets on other topics (Shepherd, Sanders, Doyle, and Shaw, 2015; Hewis 2015; Burch, Frederick, and Pegoraro, 2015). A benefit that qualitative analysis such as thematic analysis offered among the studies included here, is that there was more depth and richness in the results that were reported. There is also a lack of empirical studies which connect the results of a study to information theory and health-related theory. The majority of health studies on Twitter analysed a single topic (Chew and Eysenbach, 2010; Scanfeld, Scanfeld and Larson; 2010; Signorini, Segre and Polgreen, 2011; Salathé and Khandelwal, 2011; Bosley, Zhao, Hill, and Shofer, 2012; Robilard *et al.*, 2013; Kostkova, Szomszor, and St. Louis, 2014; Jin *et al.*, 2014; Oluwafemi, Elia, and Rolf, 2014; Young, Rivers and Lewis; 2014; Santos and Matos, 2014; Cavazos-Rehg *et al*., 2015; Kendra, Karki, Eickholt, and Gandy, 2015; Hewis, 2015; Odlum and Yoon 2015; Shepherd, Sanders, Doyle, and Shaw, 2015; Liu, Mei, Hanauer, Zheng, and Lee, 2016; Stefanidis *et al*., 2017). Examining more than one dataset related to a similar topic i.e., infectious diseases, as the present research is proposing, may allow important similarities and potentially interesting differences to emerge.

2.16 Summary

The main aim of this literature review was to review various studies on health research related to Twitter, synthesise knowledge, establish the context of the topic, and rationalise its significance. A further key objective of the review was to identify the research techniques and methodologies in the field, to place the research in a historical context, and to show awareness of current developments in the field. Previous research has highlighted the potential of Twitter for disease surveillance and early warning detection. Moreover, studies highlighted that Twitter data has the potential to provide insights into the public's views and the perceptions of the public during emerging news events, such as infectious disease outbreaks. Twitter may also act as a key communication tool for releasing information during an outbreak and as a means of keeping the public up-to-date with developments. It is therefore important to study the content on the platform in order to understand better the types of health information the general public may have access to. The literature review found that many of the studies conducted on infectious disease outbreaks were quantitative in nature and might lack qualitative insights, a potential gap in knowledge. A qualitative approach will involve a certain level of interpretation and could potentially uncover trends and conversations that would be undetected in a quantitative study. The literature review also found that there was a lack of empirical research comparing health topics. The number of different tweets retrieved in the studies varied greatly, and it was observed that qualitative studies would utilise fewer tweets in comparison to quantitative studies due to the time-consuming nature of qualitative analysis. The next chapter will provide an overview of the methodological considerations for this study. The literature review has drawn interest from researchers from a number of disciplines and will be of interest to academics within the fields of sociology, psychology and those who work across health sectors.

Chapter 3 Methodology

3.1 Introduction

The previous chapter provided an overview of the key literature on Twitter in relation to health topics and outlined a number of social cognition models and concepts related to information theory (section 2.4 and 2.5). This chapter will introduce the methodological considerations of this study, which relate to the research philosophy, the results from a pilot study, ethical considerations, the methods and tools of data retrieval and analysis, and will discuss alternative methodologies of data analysis that were considered for potential application, but were discarded in favour of thematic analysis, which was better positioned to address the research questions.

3.2 Research Paradigms

This section will outline a number of paradigms that are commonly utilised in the social sciences to then focus on the one selected for this study. Guba and Lincoln (1998) stated that a paradigm can be thought of as “a set of *basic beliefs* (or metaphysics) that deals with ultimate or first principles” (p.107). Kuhn (1962) noted that a paradigm is “the set of common beliefs and agreements shared between scientists about how problems should be understood and addressed” (p.45). Pickard (2007) defined a paradigm as a “world view that is accepted by members of a particular scientific discipline which guides the subject of the research, the activity of the research and the nature of the research outputs’’ (p.5). Paradigms are important because they inform the approach taken by a study in relation to the ontology of a study, the epistemology, and the methodology. Table 3-1 below provides an overview of the different research paradigms relevant to this study.

Table 3-1 Summary of paradigms relevant to this study

Paradigm	key beliefs
Positivist	• There is a single reality which can be measured. • The researcher is considered as an "objective observer", independent to what is being observed and who reports on observations (Bryman, 2004; Pickard, 2007). • The approach taken may be deductive and key constructs relate to measurement and objectivity. • More closely associated with quantitative methods of data collection and analysis.
Interpretivist / Contructivist	• Interpretivists note that "reality is socially constructed" (Mertens, 2005, p.12). • Interpretivism is dependent on "participants views of the situation that is studied" (Creswell, 2003, p.8). • The approach taken may be inductive and qualitative data collection and analysis may be applied.
Pragmatic	• Pragmatism does not commit to a single world-view, and tends to focus on the 'what' and 'how' of the research problem (Creswell, 2003, p.11). • The research question is central and it involves any method which can be used to better understand the problem (Creswell, 2003, p.11). • Often associated with a mixed methods approach.

This research utilised a pragmatic research paradigm which consisted of a mixed methods approach.

Mackenzie and Knipe (2006) wrote that the quantitative and qualitative debate, i.e. that one method is better than another, has been discussed for a half a century. Punch (2013) wrote that we can think of quantitative data as are "in the form of numbers (or measurements)" (p.1) and qualitative research as "data that are not in the form of numbers (most of the time, though not always), this means words" (p.1). Related to this debate, Mackenzie and Knipe (2006) quoted an extract from a 1946 paper:

> “Social scientists have come to abandon the spurious choice between qualitative and quantitative data; they are concerned rather with that combination of both which makes use of the most valuable features of each. The problem becomes one of determining at which points he [sic] should adopt the one, and at which the other, approach” (Merton and Kendall, 1946, pp.556-557)

Since at least 1946 then, researchers have argued that those within the social sciences have come to question the qualitative and quantitative division (Mackenzie and Knipe, 2006). Pragmatism does not commit to any one system of philosophy or reality, and pragmatic researchers focus on the ‘what’ and ‘how’ of a research problem (Creswell, 2003; Mackenzie and Knipe, 2006). As the quote above indicates, the research project is not restricted to either qualitative nor quantitative approach, but rather considers the research question as central and uses a combination of methods in order to answer the research question. Consequently, in pragmatic research, the data collection and methods of analysis are chosen on the basis of whether they are likely to provide insights into the research question (Mackenzie and Knipe, 2006). In light of the above considerations, a pragmatic approach was best suited for the exploratory nature of this research. This is because the research questions that were selected for this present study were exploratory in nature.

The qualitative approach of thematic analysis was selected for this study; this will involve interpreting and developing meaning from the data. The data analysis phase of this study, therefore, involved elements of interpretivism. Mason (2002) wrote that:

> “An interpretive reading will involve you in constructing or documenting a version of what you think the data mean or represent, or what you can infer from them” (p.149)

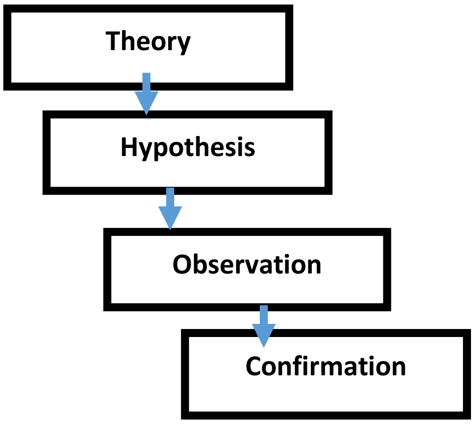
Different researchers may interpret the content of tweets differently. Hence, to provide consistency and robustness, a number of reliability measures were utilised such as intercoder-reliability, test-retest reliability, as outlined in section 3.8.

3.3 Deductive and Inductive Reasoning

There are two broad methods of reasoning which are known as the deductive and inductive approaches (Bryman, 2008; Trochim, Donnelly, and Arora, 2014). Deductive reasoning goes from the general to the more precise, and is informally termed as a top-down approach (Figure

3-1) (Bryman, 2008; Trochim, Donnelly, and Arora, 2014). A researcher may begin with an overarching theory about a topic or interest, and narrow it down to a more specific hypothesis, which can then be tested (Bryman, 2008). It is also possible to focus further when collecting observations (data) (Trochim, Donnelly, and Arora, 2014). This gives the researcher the ability to test hypotheses, and allows for the confirmation or rejection of the original theory. Figure 3-1 below is a diagrammatic representation of the process of deductive reasoning.

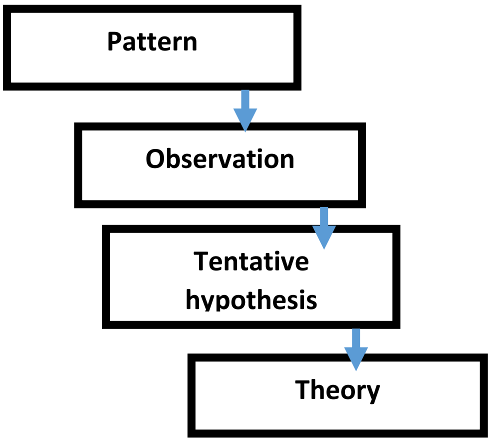
Figure 3-1 Process of deductive reasoning



Adapted from: *Trochim, Donnelly, Arora (2014)*

Inductive reasoning is distinct from the deductive approach because it moves from the more precise to generalisations and theories which are broader in nature (Figure 3-2); this is also informally known as a bottom-up approach (Bryman, 2008; Trochim, Donnelly and Arora, 2014). In inductive reasoning, the research may start with observations which are very specific, which allows for the detection of potential patterns in order to develop initial hypotheses that can be further explored, and finally this approach will outline general conclusions and theories (Bryman, 2008; Trochim, Donnelly and Arora, 2014). Inductive reasoning is more open-ended and exploratory, especially during its initial phases, whereas deductive reasoning tends to have a narrow focus and is based on testing a hypothesis (Bryman, 2008; Trochim, Donnelly, and Arora, 2014). Most social science research involves both inductive and deductive reasoning processes at some point in the project (Trochim, Donnelly, and Arora, 2014).

Figure 3-2 Inductive reasoning



Adapted from: *Trochim, Donnelly, Arora (2014)*

The present study looked to explore and understand the information that was shared on Twitter during the Ebola outbreak of 2014, the swine flu outbreak of 2009 and as well as the 2017 Zika outbreak. This research project is, therefore, inductive in nature as no theory or hypothesis was being tested initially. The Health Belief Model and Information Theory were applied when interpreting the results. A justification for utilising the Health Belief Model can be found in section 2.5.3 in the Literature Review, and justification for applying Information Theory to the results can be found in section 2.4.

3.4 Research Strategies and Research Design

There are a number of research designs associated with either the quantitative or qualitative research methodologies. These are summarised in Table 3-2 from Bryman (2008).

Table 3-2 Overview of research strategies - adapted from: Bryman, A. (2008).

Research Design	Research strategy	
	Quantitative	**Qualitative**
Experimental	Quantitative contrasts between experimental control groups in regards to dependent variables.	Bryman (2008) noted that experimental qualitative research does not have a typical form, but that in previous research data had been collected using a quasi-experimental research design.
Cross-sectional	Associated with survey research or within structured observation at a distinct point in time, e.g. content analysis on documents.	Qualitative interview or focus groups at individual time point e.g., content analysis using qualitative methods on a series of documents based on a single time point.
Longitudinal	Survey based research which is takes place more than once, i.e., in panel and cohort studies. For example, similar statistical analyses performed on a set of data from two time periods.	Ethnographic research over a significant time period, qualitative interviews that take place more than once, or qualitative content analysis for same documents which take place at two points in time.
Case Study	Studying a single case e.g. a single case is studied i.e., organisation, utilising quantitative methods.	In-depth study utilising ethnographic and/or qualitative methods of a single case i.e. intensive study by ethnographic or qualitative interviewing of a single case, e.g. organisation.
Comparative	Survey research where there is a comparison between two or more cases, i.e., which may occur in cross cultural research.	Ethnographic or qualitative interview on single and/or multiple cases.

This study employs a case study research design and examines a two-day period related to the swine flu outbreak in 2009 and a two-day period on the Ebola outbreak in 2014 (the sampling strategy for the two cases can be located in section 3.12). An additional case is examined realting to a two-day period of Twitter activity on the Zika outbreak. Previous research on the use of Twitter for health (Chew and Eysenbach, 2010; Scanfeld, Scanfeld and Larson, 2010; Signorini, Segre and Polgreen, 2011; Bosley, Zhao, Hill, and Shofer, 2012; Robilard *et al.*, 2013; Santos and Matos, 2014; Jin et al, 2014; Oluwafemi, Elia, and Rolf, 2014; Kostkova, Szomszor, and St. Louis, 2014) used some form of case study approach, but only one study explicitly mentioned the use of a well-defined case study (Gupta, Joshi, and Kumaraguru, 2012). Hartley (1994) noted that case studies are tailor made for areas of research that are not well understood because of the flexibility provided by the approach.

A case study approach traditionally involves the observation of a particular contemporary phenomenon in 'its real life context using multiple sources of evidence' (Robson, 2002, p.178). Saunders, Lewis and Thornhill (2009) noted that this strategy is useful if the aim is to obtain an in-depth understanding around the background of a study and that the case study strategy has the ability to generate answers to questions such *as 'why?', 'what?',* and *'how?'*. Moreover, researchers have noted that case studies are best applied to exploratory research because they are capable of providing insights into current events over which the researcher will have no control (Yin, 1994).

There are various research strategies that can be utilised when conducting research that may belong to either the deductive or inductive approach outlined in section 3.3. However, Saunders, Lewis, and Thornhill (2009) noted that research strategies should not be thought of as mutually exclusive (Saunders, Lewis, and Thornhill, 2009; Bryman, 2008).

In the literature review, many studies were 1) identified a topic of interest, 2) sourced data, 3) analysed the data, and 4) developed conclusions based on the analyses. As noted by Nelson, and Staggers (2013), research conducted on Twitter can be divided into three areas: description of the content on social media sites (i.e. analysing the content), use of social media, and potential use of social networks for research. The current study: 1) reviewed literature and outlined the theoretical position of the study, 2) identified an area of interest on Twitter, 3) sourced Twitter data, 4) selected a method of analysing tweets, 5) applied the Health Belief Model and Information Theory to interpret the results, and finally 6) drew conclusions in relation to theory.

This study sought to understand the results in relation to information theory, health theory as well as a number of sociological concepts. Previous evidence-based research, which used a case study approach in relation to Twitter, has lacked a theoretical basis (Chew and Eysenbach, 2010; Scanfeld, Scanfeld and Larson, 2010; Signorini, Segre and Polgreen, 2011; Bosley, Zhao, Hill, and Shofer, 2012; Robilard *et al*., 2013; Kostkova, Szomszor, and St. Louis, 2014; Santos and Matos, 2014; Jin *et al*., 2014; Oluwafemi, Elia, and Rolf, 2014). It has been noted that research which is conducted alongside theory is "immediate, insightful, and applicable in practice" (Reeves, Albert, Kuper, and Hodges, 2008, p.634). This is not to say that research which lacks theory is inferior, but rather that studies which align themselves with theory may appear to be more insightful when the results are interpreted. Other scholars such as Lewin (1951) has also noted that "there's nothing so practical as good theory" (p.169), as theory has the potential to allow for more practical conclusions to be drawn. Therefore, one of the potential novel contributions to knowledge that the present study offers is that it is the first research to connect information and health theory to the analysis of Twitter data on two major infectious disease outbreaks this is discussed further in section 7.6.

3.5 Overview of Research Dimensions

As previously outlined in section 3.2, the research project operated under the pragmatic paradigm and utilised a mixed methods approach in order to provide answers to the research questions. The research method that was employed in this study was exploratory; however, it involved a considerable in-depth qualitative analysis (thematic analysis) which involved elements of Interpretivism, and included a quantitative element (i.e. the counting of numbers in the themes). The in-depth qualitative methodology was used to analyse tweets and to develop a number of themes and sub-themes, and a number of quantitative analyses were conducted to complement the qualitative analyses.

3.6 Overview of Data Sources Used in the Study

Table 3-3 below provides an overview of the data sources used within this study, which included publication databases (for the literature review), Google search, conference publications, and Twitter (for the review of software and the analysis of the tweets).

Table 3-3 Overview of data sources used in the study

Research phase	**Data source**
Literature review	Relevant publications in electronic databases
Comprehensive review of software	Google search, conference publications , Twitter
Analysis of tweets	Twitter

3.6.1 Twitter as Source of Data

As noted in the previous section, the data source that was analysed in this study was Twitter data. Twitter has an estimated 328 million monthly active users (Twitter Q1 2017 Company Metrics, 2017). More generally, the number of adult Internet users is increasing steadily (Duggan, Ellison, Lampe, Lenhart, and Madden, 2014), and in 2016, 24% of adult Internet users worldwide used Twitter (18% in 2013), accounting for 19% of the entire adult population (Pew

Research Centre, 2016). Twitter is most popular with people under 50 years of age and who are college educated (Pew Research Centre, 2016). Obviously, not everyone uses Twitter as their social media platform, and not all Twitter users tweet about health topics. Therefore, Twitter data does not represent all members of the public nor does it represent all Internet users (Nelson and Staggers, 2013). However, Twitter has become a popular platform for research related to public health (Sinnenberg, 2017), and has been used in a number of studies to track health trends, as well as behaviour patterns related to illnesses and medical condition e.g. suicide risk (Sueki, 2014).

When conducting research on Twitter it is possible to retrieve data via the use of keywords and/or hashtags. However, the keywords that are used to retrieve data may not be capable of gathering all the data which is generated on a topic, for instance, if a research project sought to capture mentions of swine flu using the keyword 'swine flu' this could exclude people who may refer to the virus without mentioning it within their tweets. Additionally, there may be content posted on social media using keywords containing 'swine flu' which could be non-relevant such as spam-based tweets. This is one of the limitations of retrieving data on social media via keywords and hashtags.

This section has outlined the use of Twitter as a data source as well as limitations of retrieving data via keywords, and the next section will outline some of the methods of retrieving this data by making use of Twitter's Application Programming Interfaces (APIs).

3.7 Overview of Data Analysis Techniques

This present study made use of content analysis which was utilised in the pilot study, thematic analysis which was utilised in the main study, and social network analysis which was used in order to identify influential Twitter users. These data analysis techniques are described below.

3.7.1 Content Analysis

Content analysis is a method which allows researchers to construct inferences from written work (other material may also be counted), which are replicable and valid (Krippendorff, 2004). The method can be used to gain new insights into a particular topic. Krippendorff (2004) noted that seven elements within written text can be counted such as words, themes, characters, paragraphs, items, concepts and semantics. Content analysis is conducted via coding frames in order to organise data and identify findings (Krippendorff, 2004). In the pilot study described in section 3.10, content analysis was applied to better understand tweets related to Ebola. In content analysis, researchers may decide to utilise existing code frames developed by researchers studying a similar topic, or develop their own code frames.

3.7.2 Thematic Analysis

As will be outlined in section 3.10, the pilot study, it was found that the qualitative insights that can be drawn from content analysis are limited because it limits the coding to a number of predefined themes and subthemes and can be considered a quantitative methodology by some researchers (Bengtsson, 2016). Moreover, as the research questions evolved over the course of the research project, it became apparent that thematic analysis was more appropriate for gaining in-depth insights and addressing the research questions of this study. Braun and Clarke (2006) suggested that thematic analysis:

> "…is a method for identifying, analysing and reporting patterns (themes) within data. It minimally organizes and describes your data set in (rich) detail. However, it frequently goes further than this, and interprets various aspects of the research topic" (p.79).

Thematic analysis was selected because it is a qualitative research method that provides rich and detailed results that may be lacking in quantitative research methods (Steckler, McLeroy, Goodman, Bird, McCormick, 1992). Moreover, the literature review (section 2.13) highlighted a lack of empirical studies that have used an in-depth qualitative approach to analyse Twitter content related to infectious disease outbreaks.

Thematic analysis for this research consisted of six phases, as advocated by Braun and Clarke (2006). The first phase involved becoming familiar with the data. The second stage consisted of generating initial codes from the data, and this was achieved using the qualitative analysis programme NVivo. The third phase involved searching for themes, and phase four involved the review of identified themes. The fifth stage involved defining and naming themes (undertaken using Microsoft Excel). Phase six involved writing reports on the results, which can be found in Chapter 4 Swine Flu (section 4.7), Chapter 5 Ebola (section 5.6), and Chapter 6 Zika (section 6.5) respectively.

3.7.3 Social Network Analysis

Social network analysis can be used to examine the relationships between entities, for example, Twitter users (Wasserman and Katherine, 1994). Social network analysis has the potential to examine the users that were most influential during an outbreak. The study utilised social network analysis in Chapter 4 Swine Flu section 4.14 , and Chapter 5 Ebola section 5.13. The two network metrics that were used to calculate influence were InDegree, and OutDegree. Wasserman and Katherine (1994) provide the example of trade networks to illustrate this concept. That is, if a country exports many goods, then they have a high OutDegree, and if they import many goods then they have a high InDegree. On social media platforms this means that users who tweet about a topic frequently will have a high OutDegree, and those users who are mentioned most by others will have high InDegree.

3.7.4 Sentiment Analysis

Sentiment analysis can place tweets into positive and negative categories in large volumes (Pang and Lee, 2008). Sentiment analysis is a popular method for analysing social media data such as Twitter (Pak and Paroubek, 2010; Kouloumpis, Wilson, and Moore, 2011). Although sentiment analysis is capable of analysing large volumes of tweets in bulk, questions may arise over its accuracy (Pozzi, Fersini, Messina, and Liu, 2016) and there will be limited depth to the data that is analysed and hence it was decided not to employ it in this study.

3.7.5 Machine Learning

Machine learning can be used to automatically place tweets into categories based on a dataset that has been trained (Pennacchiotti, and Popescu, 2011). For example, a researcher may code a subset of tweets in order for a machine learning classifier to code the bulk of the dataset.

This can allow researchers to code large volumes of Twitter data with ease. Although this method was also considered, it was found that the qualitative insights that could be drawn were limited, as the output of machine learning would only specify how many times a tweet occurred in relation to a theme.

3.8 Quality of Research

When undertaking research on social media, it is important to transparently document the sorts of Application Programming Interface (API) used, the dates, and the types of data retrieved (Vis, 2013). This is one measure of assuring the quality of social media research. This present study has carefully documented the procedures of how the data was retrieved as well as how the data was filtered and analysed which ensures other researchers could replicate this present study. Previous studies published in peer-reviewed journal articles utilising the method of thematic analysis on Twitter data have not performed additional reliability measures (Heaivilin, Gerbert, Page, and Gibbs, 2011; Shepherd, Sanders, Doyle, and Shaw, 2015; Hewis, 2015; Burch, Frederick, Pegoraro, 2015). This present study, however, utilised two techniques of assessing the reliability of the coding process. The first technique is known as Intercoder reliability, which is described in section 3.7.1, and the second technique is known as test-retest reliability, as described in section 3.7.2.

3.8.1 Intercoder Reliability

A technique that can be applied to content analysis and thematic analysis is intercoder reliability. Intercoder reliability refers to how much two independent coders will agree when coding content based on the same coding scheme (Cho, 2008). The procedure for undertaking intercoder reliability consists of a second independent coder coding the data in order to compare their results with those of the original coder. Neuendorf (2002) noted that researchers would rarely need to code more than 300 units when producing intercoder reliability as the statistics would not change as the number increased. Moreover, in terms of the feasibility and practicality of a PhD study, it would not have been possible to source a second coder to code the entire dataset of tweets.

3.8.2 Test-Retest Reliability

Test-retest reliability is calculated in a similar manner to intercoder reliability, however, it involves using the same researcher who coded the initial dataset to code a subset of tweets from the same data after suitable length of time has passed. This form of reliability rose to prominence in the for experimental work in the fields of psychology and sociology in order to assess the level of variation if the experiment was to be repeated (Guttman, 1945).

3.9 Application Programming Interfaces

An application programming interface (API) is a set of programming instructions or standards for accessing web-based software applications and to access web-based tools (Roos, 2015). Software companies may release their API to the public to allow software designers to build products powered by their service (Roos, 2015). Amazon, for instance, has released an API which allows other web developers to access Amazon's product information. Using the Amazon API, a third party website can provide links to Amazon products with their updated prices and with a "buy it now" option (Roos, 2015). In the case of Twitter, allowing third party developers a partial access to the Twitter API implies that they can create programs to incorporate Twitter's services such as retrieving Twitter data (Strickland and Chandler, 2015). Knowledge of the Twitter API will inform the decision of selecting the appropriate Twitter API for gathering Twitter data.

The majority of research on Twitter has relied on using two particular APIs, the Streaming API and the Search API (Gaffney and Puschmann, 2014). This still appears to be the case three years on, potentially because of the high cost of purchasing Twitter data. The Streaming API is utlised most widely in quantitative research and the majority of quantitative research on Twitter data (Gaffney and Puschmann, 2014). The Streaming API is known as a push based service, which means that the data is said to be captured live. This might require researchers to build tools in order to maintain the connection to the server (Gaffney and Puschmann, 2014). The Streaming API is provided through three bandwidths: 'Spritzer', 'Garden-hose', and 'Firehose', which can deliver 1%, 10% and 100% of tweets respectively, over a given time period (Gaffney and Puschmann, 2014). The Spritzer bandwidth can be used for academic research and any ordinary Twitter account will be given access to the Spritzer bandwidth. The Garden-hose is supplied to users who had compelling reasons for increased access and was also supplied to businesses or licensed re-sellers. The 1% cap for the Spritzer bandwidth is triggered if the tweets about an event exceed 1% of all traffic on the platform (Gaffney and Puschmann, 2014). In a conference hashtag, for example, the spritzer bandwidth would be sufficient for capturing all the tweets, because only in exceptional cases would all tweets about an event exceed 1% on Twitter, for example, during a major disaster, such as an Earthquake, or a sports event, such as a football final. Table 3-4, below, summarises the different bandwidths and % of tweets that are returned.

Table 3-4 The different bandwidths for the streaming API return and the % of tweets returned

	Bandwidth	**% of tweets returned**
Streaming API	Spritzer	1%
	Garden-hose	10%
	Firehose	100%

The Search API is similar to the Search feature available in Twitter web clients and mobile devices (Twitter Developers Blog, 2014). The Search API is focused on relevance rather than completeness, and therefore some tweets and users will be missing from the results (Twitter Developers Blog, 2014). Thus, early work examining Twitter (Gaffney, 2010) was based on the Search API (Gaffney and Puschmann, 2013). Twitter has discouraged the use of the Search API, because it is not feasible to maintain and may be discontinued in the future (Gaffney and Puschmann, 2013). However, at the time of writing (Q4 2017) the Search API is still available to use.

Gaffney and Puschmann (2013) suggested that, although it may be possible to collect historical data from the Search API, this method of data collection is severely limited, because search results may no longer appear after a week. In addition, there is no information regarding the completeness of the data. The Search API provides users with methods that are different to the Streaming API. The Search API is pull-based, meaning that a researcher or a program can request data from the server and the server will return the data. The Search API is rate limited, which makes it difficult for researchers to collect Twitter data in a timely manner. Rate limiting occurs when demand is high on Twitter, or if a particular IP address has been retrieving Twitter data. The Search API rate limits have changed from version 1.0 to 1.1 of the REST API, rate limits in 1.1 of the API are allocated into 15-minute intervals (previously 60-minute intervals).

In addition, in version 1.1 of the Search API, users are not able to make unauthenticated calls, that is to say, users must register with Twitter to receive a special username and password known as OAuth authentication (Twitter Developers Blog, 2014). Vis (2013) also noted that most academics have to rely on public APIs, which provide free data (a limited number of tweets), as it is not feasible for researchers to obtain and analyse Firehose data (all the available tweets) for financial reasons.

González-Bailón *et al*. (2014) sampled Twitter activity around political protests on the Spanish indignados movement from 30th of April 2012 to 30th of May 2013. They collected two independent samples related to messages on the protests. The first sample was collected using the search API from the UK using five of the most popularly used hashtags. The second sample was collected from Spain, using the streaming API with 70 hashtags. They found that the structure of the two samples was considerably affected by the API used and the ammount of hashtags that were used to retrieve the data. The implication of this is that the type of API used and the number of hashtags will affect the sample that is obtained from Twitter.

Gerlitz and Rieder (2013) analysed 1% of all publically available tweets in order to reflect on the different methods of collecting data from Twitter; in particular, the authors were interested in using a random sampling technique, which is made possible via the Twitter Streaming API. Gerlitz and Rieder (2013) suggested that the Streaming API is unique compared to REST based implementations. The Streaming API requires a constant connection with the Twitter servers and the tweets are sent in almost real time. Gerlitz and Rieder (2013) highlighted that there is a 'statuses/Firehose' endpoint, which provides all available tweets to a selected client, whereas the 'statuses/sample' endpoint provides a small random sample to a client. Gerlitz and Rieder (2013) indicated that Twitter's senior partner engineer, Taylor Singletary, suggested in a forum post that the sample stream would be a random sample of all tweets that were available on the platform. Gerlitz and Rieder (2013) noted that, with regard to standard sampling, the 1% statuses end point would provide a sample that was representative (of the Firehose), and that if it is not possible to obtain Firehose data, then random sampling via the Streaming API could act as a baseline for research.

Morstatter, Pfeffer, Liu and Carley (2013) compared the Streaming API with the Firehouse API in order to determine whether the Streaming API is representative of all Twitter activity. They found that, when estimating the top number of hashtags in the Streaming API, the data may be distorted when the number of captured hashtags was low (although it would improve when the number increased). Consequently, the topical distribution of tweets via the Streaming API became more representative as the amount of collected data increased. Morstatter, Pfeffer, Liu and Carley (2013) reported that the Streaming API will, at most, return 1% of all tweets. This is because, when all tweets that match a particular query exceed 1% of all tweets, Twitter begins to sample data, which is returned to a user (Twitter Developers Blog, 2014). Morstatter, Pfeffer, Liu, and Carley (2013) reported that one method to overcome the 1% limit is to use the Firehose. The Firehose provides a complete 100% access to all available tweets on Twitter. However, there are two critical limitations of Firehose data: firstly, the cost of obtaining Firehose and secondly, the cost of maintaining the data (device storage and servers, etc.) are not feasible for most researchers.

Borra and Rieder (2014) indicated that people working on Twitter's in-house research projects would have complete access to all available tweets. Thus, these researchers would not have to worry about the completeness of data sets, data accessibility, and technical infrastructure. However, academic ethical standards would still apply to them as well as legal considerations, i.e. not sharing data would still apply to them as they received Twitter data through the goodwill of Twitter and it could also be difficult to assess their academic independence. The tools and techniques employed by in-house researchers are owned by Twitter, which makes it difficult for other academics to use them. It is possible to purchase data via licensed resellers such as Gnip and DataSift for the full archive of tweets; however, this is not feasible for most researchers, especially due to the subscription services, which are too expensive for most academic research groups. This in turn complicates the process of peer review, if the reviewer is technically unable to replicate results.

This section has outlined the concept of an API, and the implications that this might have when undertaking research on Twitter. The next section outlines an experiment that was undertaken as part of the PhD, in order to learn how to retrieve Twitter data, and to also compare the number of tweets that different social media software tools were capable of gathering.

3.10 Comparison of Twitter Application Programming Interfaces (APIs)

In the first year of this PhD, a number of social media software tools were used to retrieve data on Ebola in order to ascertain the amount of tweets they could retrieve. This is because different software applications will utilise different APIs, which provide different levels of data access, and the applications may retrieve Twitter data in varying time intervals. Moreover, this exercise also provided the researcher with an opportunity to learn how to retrieve Twitter data. The time period that data were retrieved from was 03/01/2015 to 05/01/2015, and this particular time interval was selected in order to be consistent across the three software tools. The Ebola epidemic was ongoing during the first year of this study, and because data on Ebola was found to be relevant to this study, this is the topic on which data were gathered for the subsequent comparison study. Table 3-5 displays the software and API that were used to gather Twitter data, as well as the amount of tweets gathered for the keyword ‘Ebola’. Thelwall (2015) notes that assessing the comprehensiveness of data used from Twitter can be considered good practice.

Table 3-5 Number of tweets gathered using different tools and APIs

Tool	API	No. tweets
DiscoverText	Firehose API	195,713
Mozdeh	Search API	155,086
Chorus	Search API	145,348

Chorus (n.d.) is a free open source software tool that uses the Search API and it was found that using this tool it was possible to gather 145,348 tweets, which are up to 74.27% of Firehose data. The next software tool that was used to retrieve tweets was Mozdeh (n.d.), and it was found that this software was capable of retrieving 79.24% of Firehose data. Exact duplicate tweets were removed in Mozdeh to ensure only single tweets were examined. These percentages were calculated by using the number of tweets retrieved from the Firehose (195,713 tweets, i.e. a 100%) and calculating a percentage out of the total that were retrieved. Previous research has reported that the Search API is capable of gathering between 1% and 40%, and the findings of this experiment found that it was significantly more than the 1% to 40% upper limit reported in the literature (Morstatter, Pfeffer, Liu, and Carley 2013) for the query ‘Ebola’.

At the time the study was published, no previous empirical research had compared the Search API to the Firehose API via different software applications. These results have been disseminated as a conference abstract (Ahmed and Bath, 2015b) and the analyses presented in this section will be developed for further publication. This method of comparing the number of tweets may be possible by utilising services such as the Quick Trends explorer in Visibrain (n.d.) which provides the number of tweets sent on a topic over the previous month. Thelwall (2015) has also proposed an alternative four-step method in assessing the coverage of Twitter data via the Search API without requiring access to historical Twitter data and/or the Quick Trends Explorer in Visibrain (n.d.).

A pilot study was conducted before undertaking the analysis of the larger PhD study, which informed methodological decisions and is outlined in the next section.

3.11 Pilot Study on World Autism Awareness Day

The original research plan of this present study was to examine a number of health topics, and was later revised in order to focus specifically on infectious disease outbreaks. A by-product of this was a study that investigated whether social media platforms such as Twitter had the potential to raise the awareness of the public during health campaigns. Mozdeh (n.d) was utilised in order to gather 2,315,283 tweets over a two-month period monitoring 23 keywords and hashtags in order to assess the successfulness of World Autism Awareness Day (WAAD). In order to determine whether the campaign was a success, a framework was developed which set out a number of criteria that should be met for a campaign to be successful which were as followed:

- The volume of tweets should increase
- There should be an increase in sentiment
- There should be an isolates group
- A significant proportion of tweets should relate to the campaign

It was found that WAAD was successful in raising awareness on Twitter because each of the criteria listed above were successfully met. The pilot study has been accepted as a preliminary paper into the iConference (Ahmed, Bath, Sbaffi, and Demartini, 2018) and will be published in the *Lecture Notes of Computer Science.* This study helped the researcher through the research journey and a number of skills were developed.

3.12 Pilot Study on Ebola Tweets

This section presents the results of the experimental case study that was conducted during the initial stages of this project. A pilot study in social science allows the imperfect first attempts at answering a research question (Martin, 2007).

3.12.1 Selection of Pilot Study Data

The pilot study concentrated on the date the 30th September 2014, as it is the first time in history that the Ebola Virus was diagnosed outside West Africa (WHO, 2015). In total, for this 24-hour period, there were 470,184 tweets, which were retrieved via the Firehose API (that is to say *all* of the available tweets). When Ebola reached epidemic status on August 8th (33 weeks into the outbreak), the WHO declared the epidemic to be a Public Health Emergency of International Concern (PHEIC). The PHEIC is an instrument of the International Health Regulations (IHR), which is a legally binding agreement among 196 countries on containment of major international health threats (Briand *et al.*, 2014).

3.12.2 Data Extraction Strategy

Tweets from the first 20 hours of the dataset were separated, and only the tweets up until the time of 8pm were extracted (leaving 80,538 tweets). This is because after 8pm, an increase in re-tweeted content and spam was observed. Out of these remaining 80,538 tweets, each Twitter hour was reported in a separate sheet in Microsoft Excel (12am to 7.59pm) and these tweets were randomised using the =RAND() feature in Microsoft Excel. From this sample, 25 random tweets (not including retweets) from each hour (19 hours in total) were separated for a total of 475 tweets.

3.12.3 Data Inclusion and Exclusion Strategy

The approach taken here was similar to that used in previous research (i.e. Chew and Eysenbach, 2010), in that the pilot study did not examine retweets as there is a risk of only examining popular content on the platform (Chew and Eysenbach, 2010). This also reduces the chance of analysing spam, i.e. tweets that are not related to the subject matter of interest (Chew and Eysenbach, 2010). It was decided that tweets with the '@' notation would not be examined, as these indicate a conversation between two Twitter users and it is difficult to determine the context. Due to feasibility, links, images, videos, and associated media files were not examined. For practical reasons, Tweets that were not in English were also excluded in this pilot study. A limitation of this approach is that by including only English tweets it may not offer a complete picture of the outbreak situation. A further potential limitation is that by excluding retweets it could make less visible content more visible and more visible content less visible. In order to counteract this limitation an overview of popular tweets was provided for the studies on Swine Flu and Ebola.

3.12.4 Coding Frames

As Twitter allowed only brief 140-character text updates known as 'tweets', these were counted and placed into categories using content analysis. The tweet content coding frames were adapted from Chew and Eysenbach (2010), after discussions with the supervisory team (see Appendix 1 for developed coding frame). Tweets which were resource based and/or which were spam-based were coded with a second set of codes (see Appendix 1, Qualifier Categories).

3.12.5 Intercoder Reliability Process

In order to support the reliability of the results presented here, a second coder was trained to code 10% of each sample, in order to produce an intercoder reliability percentage, and associated Scott's pi (*π)* and Cohen's kappa (*κ)* statistics. Intercoder reliability is the amount to which two coders will agree on the coding of the content of interest using the same coding scheme (Cho, 2008). This was calculated using ReCal (Freelon, 2010). An instruction sheet (Appendix 1) and the data to be coded (a 10% sample) was provided to a coder (a researcher in health and social media research). This coder, using the instruction sheet, independently coded 10% of the data sample within Microsoft Excel. This process allowed the researcher to gain feedback on the coding frames, which then improved, based on this process. The disagreements between the coders for the first code frame included:

- Coder 1 = Ambiguous & Coder 2 = Jokes Parody
- Coder 1 = Personal Opinion & Interest & Coder 2 = Ambiguous
- Coder 1 = Spam & Coder 2 = Resource

For the second part of the code frame, the disagreements consisted of

- Coder 1 = N/A & Coder 2 = Humour
- Coder 2 = N/A & Ambiguous

These disagreements were then discussed and were found to depend upon differences in opinion between the two coders. No further action was taken because the disagreements were differences in interpretation rather than errors in coding.

3.12.6 Results of Tweet Coding

Table 3-6 shows the results of the tweets that were placed into categories using the tweet category code frame (Appendix 1).

Table 3-6 Frequency of tweet content

Category	Frequency	Percent (%)
Resource	385	81.1
Personal opinion and interest	46	9.7
Jokes/parody	15	3.2
Ambiguous	14	2.9
Marketing	2	0.4
Spam	2	2.7
Total	464	100

As shown in Table 3-6, 81.1% of Twitter users were sharing Ebola-related resources such as Ebola news, updates or information, 9.7% of tweets were sharing the user's personal opinion and interest in Ebola, and 3.2% were expressing humour related to the Ebola outbreak (such as jokes or parodies). Intercoder reliability percentage agreement for this data sample was 94.0%, $\pi = 0.82$, $\kappa = 0.822$, which indicates a good level of agreement (McHugh, 2012). Intercoder reliability is a standard measure of research quality, and a low level intercoder reliability may suggest weakness in coding methods. Cho (2008) wrote that the *percent agreement*, due to its simplicity and ease of use, is the single most widely used index. Table 3-7 below displays tweets that were coded according to the second code frame, i.e. all tweets apart from those that were assigned to the category of resource, or spam.

Table 3-7 Qualifier categories

Content	**Frequency**	**Percentage of tweets assigned to each category**
Humour or Sarcasm	22	4.4
Concern	14	2.8
Frustration	6	1.2
Misinformation	5	1.0
References Popular Culture	5	1.0
Question	3	0.6
Ambiguous	3	0.6
Fear	2	0.4
Relief	1	0.2

Ebola resources and spam were not placed into qualifier categories (see appendix 2). In other words, if a tweet was providing an Ebola-related resource, or was spam, as defined in the code frame, then this tweet was not placed into a qualifier category. Table 3-8 shows that of the tweets that were placed into qualifier categories, 4.4% were humorous or sarcastic, 2.8% expressed concern, 1.2% expressed frustration, and 1.0% expressed medical misinformation. The intercoder reliability percentage agreement was 96.0%, $\pi = 0.817$, and $\kappa = 0.819$, which shows good agreement (McHugh, 2012). Content analysis was limited in that it forced a particular tweet to be assigned to a category when a better method may be to generate codes and themes from a particular dataset (Joffe and Yardley, 2004). Percentages were calculated by taking the total number of tweets for each category against the total number of tweets that were analysed.

3.12.7 Discussion

The aim of this pilot study was to examine the feasibility of using content analysis as a method for analysing a corpus of tweets. A limitation of this pilot study is that only a specific day was examined, the 30th September 2014, rather than a sequence of days. Examining multiple days increases the volume of tweets that would be analysed; therefore, it was decided to examine a single time period. In addition, the pilot study has not examined web-links that are shared on the platform, images, RTs, and associated media files.

3.12.8 Conclusions

It was found that the most popular content on the platform for this pilot study on Ebola consists of news, updates, and information related to the outbreak (76.8%). A by-product of this pilot study was the development of two code frames for categorising tweets. The results of this study were presented at the iConference 2017 in Wuhan, China (Ahmed, Demartini, and Bath, 2017). This pilot study informed the development of the main study, because it was decided that a more in-depth method to analyse tweets would potentially provide richer results.

3.13 Ethical, privacy and Copyright Issues

Considerations regarding ethics preceded the design and implementation of the study and are outlined in this section. Twitter data is typically generated by people (tweets can also be generated by automated accounts). Tweeting and reviewing are actions that involve human participants and a number of possible risks needed to be addressed. Twitter data is publically accessible; however, gaining ethics approval is necessary because, personally-identifiable data are involved, and when the data are analysed, the results could draw attention to particular individuals, groups or trends. This would go beyond what the users of these platforms would expect by engaging on Twitter.

Initially, ethics approval was sought, in accordance with the University of Sheffield research ethics policy (The University of Sheffield, 2005). The University of Sheffield recently updated its ethics policy on social media research alongside changes in the Data Protection Act, and although this was published in December 2016, this present research project followed the updated policy (The University of Sheffield, 2016). The researcher was also aware of the various ethical guidelines, such as the Association of Internet Researchers (AoIR), that provide advice on such debates. Research approval was originally sought on 27/11/2014, and as the research questions of the project evolved a further ethical approval was required, and this was obtained on the 14/07/2016. Appendix 1 contains a letter which confirms ethical approval was granted from the University of Sheffield's ethics review process. This section will identify the ethical issues faced and how these were overcome.

3.13.1 To what extent is user-generated content public?

In the ethics application, it was reasoned that data on Twitter can be considered public because anyone with Internet access can view content on Twitter. There is no need to subscribe, enter a password, or pay to access the data. The research captured only tweets defined as public. Twitter informs new users in its 'terms of service' that public tweets are visible to anyone regardless of whether they have a Twitter account or not. Boynton (2014) wrote that:

> "Twitter messages are 140 characters or less. They are easy to write and easy to read. And they are by default public. Most messages can be found and read in a variety of ways. Unlike email, unlike texting, unlike messages on Facebook these messages are in the public domain." (p.79)

It is important to note that the question of whether Twitter data is public is currently an area of debate (Beninger *et al.*, 2014) and the present research took the stance that Twitter data can be considered public.

3.13.2 Copyright on Tweets

The ethics application also outlined how the use of copyrighted works would be allowed. The Copyright, Designs and Patents Act 1988 allows exceptions to the use and analysis of copyrighted work, for:

i. Non-commercial research and private study
ii. Text and data mining for non-commercial research
iii. Criticism, review and reporting current events

Therefore, this research was deemed not to infringe copyright as it falls under category of non-commercial research, so data from Twitter could be used legally in this study.

3.13.3 Potential Participants

The ethics application (Appendix 1) stated that all captured and relevant tweets would be analysed (i.e. those relating to relevant epidemics and pandemics or other health topics). These tweets could be posted by any user with a Twitter account (e.g. anyone from the general public or from organisations). It was decided that public figures with Twitter accounts would not be sampled specifically, or purposefully; however, their tweets may have been incidentally part of the data captured. It was decided that Tweets with geographical locations (geotag data) would only be analysed at an aggregate level, i.e. at the country level from where people were tweeting. It was also made clear in the application that Tweets with geographical locations would not be used to identify individual users.

3.13.4 Informed Consent

It was decided that it would not be practicable to gain informed consent to analyse the tweets, especially for the large data sets containing user-generated content (a sample of tweets may contain in excess of a hundred thousand items). However, during the analysis of the tweets it may become apparent that tweets from a user or a set of users are of particular interest. In this case, if it were necessary to quote the content of the user(s) verbatim, (e.g. for the purpose of reporting and substantiating the results, or for the user IDs to be indicated, for example, in the PhD thesis or in a publication), then informed consent would be sought. In the case of Twitter, this would have involved sending a tweet to the user with details of the study requesting permission to quote their tweet. However, it may have been difficult to obtain consent via participant information sheets and consent forms as users may not wish to reveal their email address or click on unknown links. In this situation, the ethics application made it possible to accept a tweet saying 'Yes' or similarly suggesting that quotes from the tweet could be used. In some instances, the researcher may have to gain consent via a tweet, email, etc. The researcher had a public Twitter account, which could have been used for this. The research project, although having procedures in place for quoting tweets in research as outlined above, took the ethical standpoint of not quoting tweets or disclosing non-public usernames unless with the permission of the user.

3.13.5 Potential harm to participants and Data Confidentiality

The data was kept on two secure password-protected laptops alongside a University backed-up secure research server. Individual tweets were not published without informed consent. In the case of Twitter's hashtag(s) generated by users on an outbreak, there was the possibility of identifying participants through de-identification techniques, that is, by searching for the hashtag using a search engine and locating participants. If this were to occur, the risk to the end users would be low, because the captured data would not fall under the category of a highly-sensitive topic, unless it related to the health of an individual (e.g. a person having contracted the Ebola Virus). In addition, if, hypothetically, the data were released by fault or negligence, the risk to the end users would be low as the captured data would not be considered a highly-sensitive topic.

3.13.6 Data Storage

The principal investigator (PI) had control of and acted as the custodian for the data captured from Twitter. The analysis of the data was conducted by the PI in the PI’s place of study (Information School) and his home. The data were not analysed in places deemed as public. The data were stored on two password protected laptops which were stored securely when not in use. Limited data were shared with the supervisors for marking or administrative use.

3.13.7 Summary

This section outlined why ethics approval was sought and the challenges that were faced. After ethical approval was gained the researcher continued to monitor debates and respond to any changes in guidelines appropriately. Although the project had ethics approval to quote tweets with consent, the researcher had taken the ethical standpoint of not disclosing tweets or usernames of members of the public.

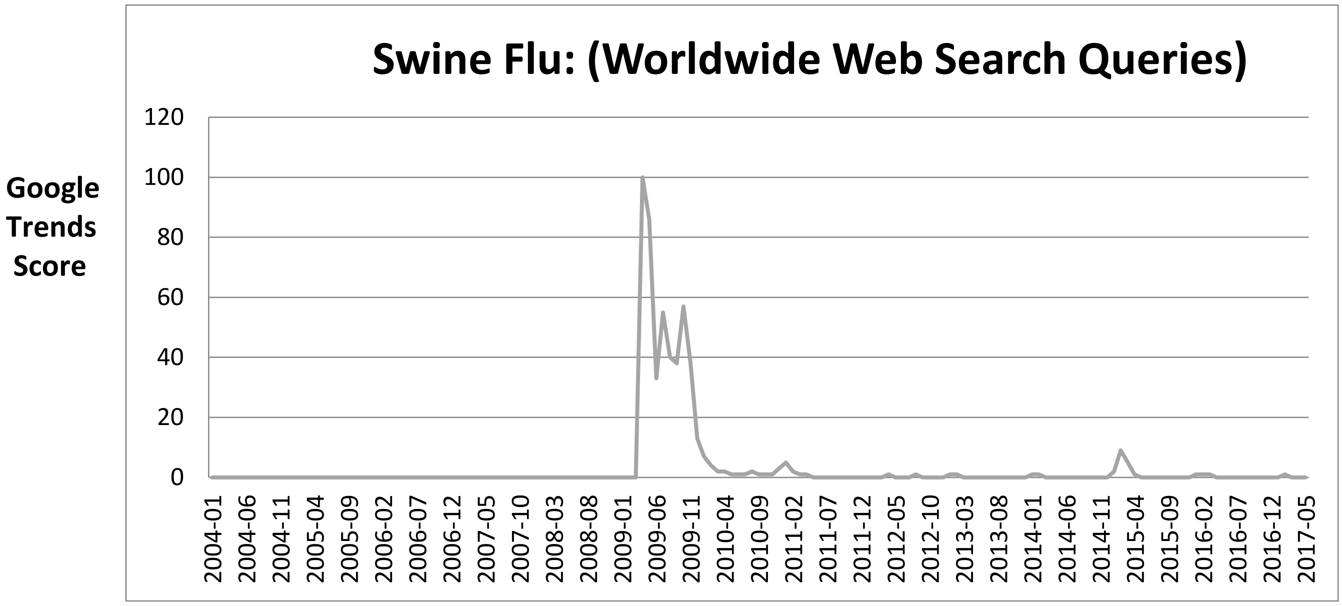
3.14 Data Gathering and Filtering Strategies

The previous section has outlined the ethical considerations for this study. This section provides a justification for the data that was gathered in this present study. Case studies were purposely selected when Google Trends was showing that there was a heightened interest in search queries for swine flu and Ebola respectively. Google Trends was used to identify peaks because, at the current time, there are no mechanisms to search all of Twitter for a particular keyword to examine peaks of interest. Google Trends was used because an increase in Web search queries is often associated with emerging news events (Carneiro and Mylonakis, 2009). Twitter is known as a platform for the detection of breaking news (Phuvipadawat and Murata, 2010); therefore, if there is an increase in Google Trends it would not be unreasonable to assume that Twitter activity would also increase around this time. The Google Trends Score is a normalised metric based on a single query. For example, two search queries could acquire a score of 100 but have different amount of searchers (Google Trends, 2017). Section 3.12.1 and section 3.12.2 outlined next will provide an overview of the decision making process and justification of selecting data to be analysed on swine flu, and Ebola respectively.

3.14.1 Selecting Swine Flu Data

The steps for selecting data related to swine flu are outlined in this section, and Figure 3-3, below, provides an overview of Worldwide Web Search queries related to swine flu. This shows that from the time period of January 2004 to May 2017 there was an increased interest in Web search queries which began in January 2009 and lasted until November 2009 (because there was an increase in the Google Trends score).

Figure 3-3 Swine flu Worldwide Web search queries

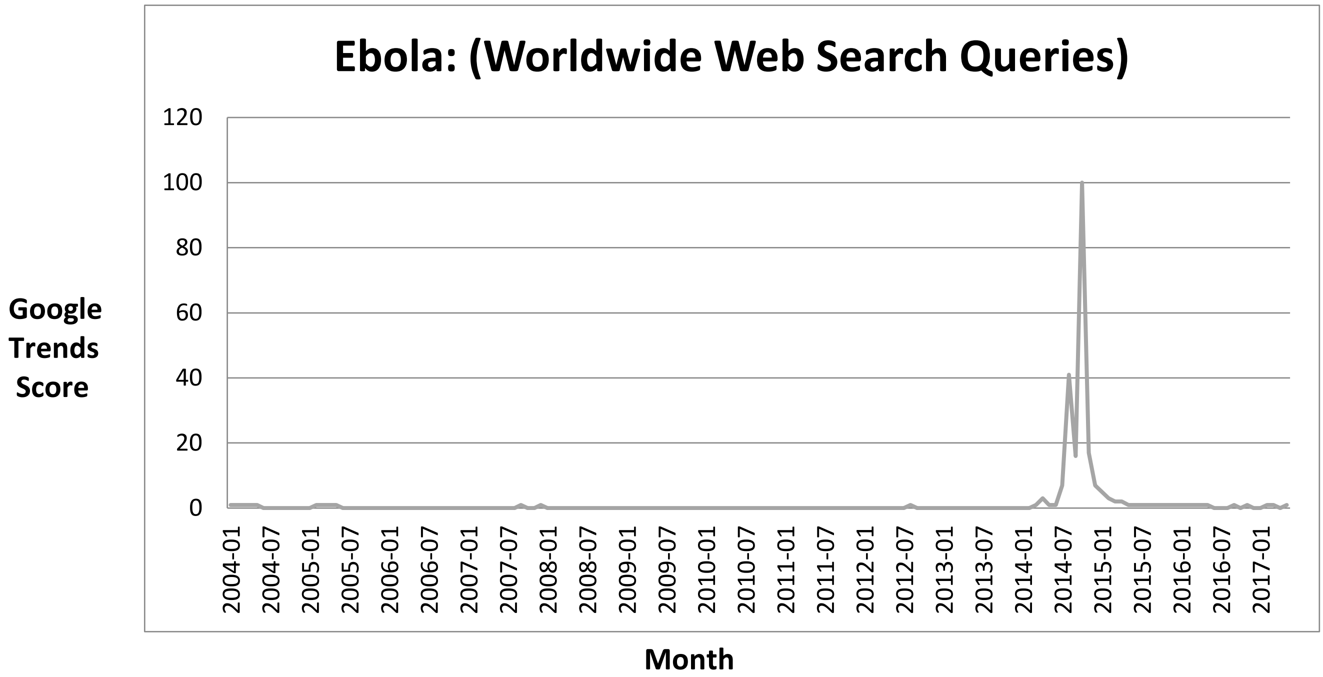


It was decided that the data for this study would be retrieved from the time period when there was a heightened interest as outlined in section 3.12. A two-day period was randomly selected from this period, and a request was sent to a licensed reseller, Visibrain (n.d.) for the retrieval of this data. The entire dataset retrieved related to swine flu consisted of 214,784 tweets posted during the two-day period of April 28th and April 29th 2009 and identified using the keywords 'swine flu', '#SwineFlu', and 'H1N1'. This time interval was selected because it falls on when there was an increased interest in the outbreak, as shown in Figure 3-3 above.

3.14.2 Selecting Ebola Data

This section focuses on the procedures taken to retrieve Twitter data on Ebola. Figure 3-4 provides an overview of Worldwide Web Search queries related to Ebola. It shows that from January 2014 to November 2014 that there was an increased interest in Web Search queries for Ebola which began around January 2014, and lasted until April 2015 (because there was an increase in the Google Trends score).

Figure 3-4 Ebola Worldwide Web search queries

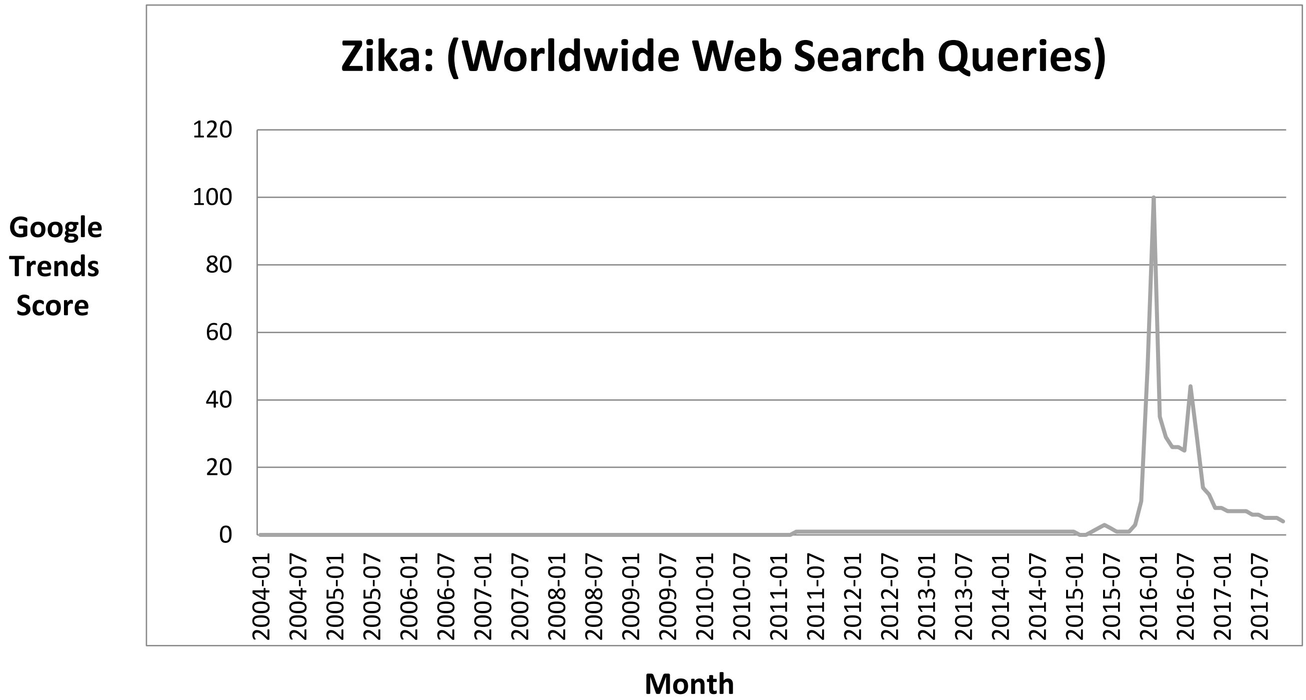


It was decided that, just as for the dataset on swine flu, a two-day period for Ebola would be randomly selected and a request was made to Visibrain (n.d.) for the retrieval of this data. The entire dataset that was retrieved relating to Ebola consisted of 181,110 tweets produced during the period of 29th and 30th September 2014 identified using the keyword 'Ebola'. This date was selected because data from Google Trends shows an increased interest around Ebola web-search queries during that time, as shown in Figure 3-4.

3.14.3 Selecting Zika Data

The original study sought to examine swine flu and Ebola and compare the results of the study to one another. Whilst this study was underway, an outbreak of Zika occurred, and because the data can be retrieved at no cost whilst an outbreak is occurring, it was decided that it would be interesting to compare the results of the analysis of swine flu and Ebola to that of Zika. The figure below shows how there was sudden interest in Web Search queries in Zika from March 2015 to July 2016 (because there an increase in the Google Trends Score).

Figure 3-5 Zika Worldwide Web search queries



The data for tweets on Zika were retrieved using the Boston University Twitter Capture Analysis Toolkit (BUTCAT) using the keyword 'Zika'. The system has access to Twitter's Gardenhose API which is at least 10% of the entire Twitter stream. 749,131 tweets were retrieved from the 2-day period 31st of January to the 1st of February 2016. This falls within a range where there was heightened interest in Zika, as shown above in Figure 3-5.

3.14.4 Justification of Data Selection

There were several reasons a two-day period for the two outbreaks was examined rather than a longer time frame. Firstly, the rationale was that this would allow for an in-depth analysis to be performed rather than a broad analysis because tweets would be of a manageable size and findings of the individual cases could be compared to one another. Research was also limited by the cost of historic Twitter data, which can become very expensive when data are retrieved over multiple days. Thirdly, in qualitative research, a key purpose is to provide findings that are rich and detailed, and which are interpreted through coding and themes (Boeije, 2009). Therefore, in order to address the research questions, which were based on providing an in-depth analysis of tweets, it was appropriate to examine a 2-day period for each of the outbreaks. However, this raises the question of whether the themes and sub-themes that were found to emerge over these days occurred at any other point throughout the outbreak.

Moreover, it can be argued that a limitation of the study is that only a 2-day period for each of the outbreaks was examined. In order to ascertain whether tweets which would express similar views across the outbreak, i.e. over a longer period of time, the study would perform a number of searchers using Twitter's Advance Search feature. Tweets were located across the outbreak periods for a number of themes, and the results of these search can be found in Chapter 4 Swine Flu, section 4.15, and Chapter 5 Ebola, section 5.12.

4.14.5 Filtering Data

In order to filter Twitter data, this study removed duplicate and near duplicate tweets, which is an important data-cleaning step necessary to avoid the possibility of only examining popular content on the platform. As pointed out by Bruns (2012), analysis which includes retweets will only serve to draw attention to the most retweeted content, therefore removing duplicates will provide a better overview of the types of content shared on the platform. Moreover, the removal of popular content has been observed previously in qualitative social media studies, for instance, Chew and Eysenbach (2010) removed retweeted tweets in order to prevent popular posts from saturating the sample, and also removed non-English tweets for feasibility purposes. The process for removing duplicate and near-duplicate tweets in this study is shown in Figure 3-6.

Figure 3-6 Removing duplicates and near duplicate tweets

All tweets		***Removing duplicates***		***Removing near-duplicates***
A1 A1 A1a	→	A1 ~~A1~~ A1a	→	A1 ~~A1 A1a~~
A2 A2 A2a		A2 ~~A2~~ A2a		A2 ~~A2 A2a~~
A3 A3 A3a		A3 ~~A3~~ A3a		A3 ~~A3 A3a~~
A4 A4 A4a		A4 ~~A4~~ A4a		A4 ~~A4 A4a~~

Jeffares (2014) wrote that:

> "The identification of near-duplicates, using DiscoverText's clustering algorithm, works by setting a threshold of what constitutes a duplicate. De-duplicating tweets is a process of clustering…A common motivation would be reductionist – to strip out near-identical tweets to speed up coding." (pp 84)

In figure 3-6, individual tweets are represented by A1, A2, A3, and A4, and variations of these tweets, for example, with different punctuation or an additional word, are represented as A1a, A2a, A3a, and A4a. The removal of duplicate tweets included the elimination of tweets that were identical and which appeared on more than one occasion. As seen from figure 7, duplicate tweets have strike-through text to indicate they have been eliminated from the dataset. However, there were still a number of tweets which were almost similar, but which may have had slight text variations, for example, in grammar or an additional word. In order to remove these from the sample, near-duplicate clusters were identified and then removed at a 60% threshold. (i.e. those tweets which were 60% similar in textual features). In figure 3-6, it can be seen that under the heading 'removing near-duplicates' A1a, A2a, A3a, and A4A have strike-through text which indicates that they have been removed from the dataset. In terms of the actual tweets that were removed below is an example of an original tweet, a duplicate tweet, and a near duplicate tweet:

Original tweet = Ebola patients are rising from the dead [URL]

Duplicate tweet = Ebola patients are rising from the dead [URL]

Near Duplicate tweet = Ebola patients rise from dead [URL]

The near duplicate tweet above has a similarity of 85% so it would be removed from the dataset. The threshold was set at 60% so all tweets that were similar 60% and above of the time were removed from the dataset. Duplicate and near-duplicate tweets were removed using DiscoverText (n.d.) in line with previous research (Jeffares, 2014).

3.14.5.1 Justification for Filtering Data

It was important to filter the respective datasets, and to remove duplicate tweets before performing an in-depth qualitative analysis, in order to avoid analysing only popular content on the platform. Furthermore, due to the volume of Twitter data it would not have been feasible to analyse tweets qualitatively without taking a random sample of tweets. Therefore, after tweets were filtered, a simple random sample of the data was taken. A simple random sample ensured an equal chance of selecting tweets and provides an unbiased representation of the entire dataset (Starnes, Yates, and Moore, 2010).

3.14.5.2 Summary of Research Approach for Filtering Swine Flu Data

For the two-day period of April 28th and April 29th 2009 selected to retrieve data on swine flu, 214,784 tweets were obtained. After the removal of duplicate tweets, the dataset was reduced to 102,852 tweets. The next stage was to locate and remove near duplicate tweets at a 60% threshold, which reduced the dataset to 76,783 tweets. Finally, a 10% random sample (7,678) of these tweets was taken and entered into NVivo in order to conduct a systematic in-depth thematic analysis. This is summarised in Table 2-8 below.

Table 3-8 Research approach for filtering swine flu data

Stage	Total
Pre-data Cleaning	214,784
Removing Exact Duplicates	102,852
Removing Duplicates at a 60% threshold	76,783
10% sample removed for analysis	7,678

3.14.5.3 Summary of Research Approach for Filtering Ebola Data

The entire dataset that was retrieved in relation to Ebola consisted of 181,110 tweets. After duplicate tweets were removed, the dataset was reduced to 102,852 tweets, and after near-duplicate were removed, a total of 56,948 tweets remained. Finally, a random sample of 10% was taken (5,695), and this dataset of tweets was entered into NVivo for coding using the data analysis technique of thematic analysis. This is summarised in Table 2-9 below.

Table 3-9 Research approach for filtering Ebola data

Stage	Total
Pre-data cleaning	181,110
Removing duplicates	102,852
Removing near duplicates at a 60% threshold	76,782
10% sample removed for analysis	5,695

3.14.5.4 Summary of Research Approach for Filtering Zika Data

The data that was retrieved on the Zika outbreak contained 749,131 tweets, considerably more than tweets on swine flu and Ebola. However, when near duplicate clusters were removed, the dataset saw the largest reduction of duplicate content. It appeared that there were a large number of news articles shared and less personal views and opinions shared on the Zika outbreak. This aligns with research that has found that there was low knowledge of

Zika in the United States (Rasmussen, Jamieson, Honein, and Petersen, 2016). This is summarised in Table 2-10 below.

Table 3-10 Research approach for filtering Zika data

Stage	Total
Pre-data cleaning	749.131
Removing Exact duplicates	76,943
Removing Near Duplicate at a 60% threshold	20,421
10% sample removed for analysis	2,042

3.15 How Data Were Analysed

3.15.1 Thematic Analysis

The method of Thematic Analysis (as outlined in section 3.7.2) was utilised in order to analyse tweets. As highlighted in section 3.12, in order to reduce the size of the dataset, a series of data cleaning techniques were applied to the dataset such as de-duplication, removal of near-duplicate clusters, and simple random sampling of tweets. After reducing the size of the dataset tweets were entered into NVivo, and figure 3.8 displays how nodes were coded. The figure displays how tweets would appear on the right hand side of the screen, and how they were highlighted and assigned a label. Each individual tweet could have been assigned multiple labels based on what was contained within the tweet. NVivo was selected because it is widely utilised in the social sciences (Paulus, Woods, Atkins, and Macklin, 2017) and supports the coding process of thematic analysis outlined by Braun and Clarke (2013). The method of analysing tweets using NVivo, as used in this study, was endorsed by QSR international, which is the organisation that founded and maintains the application, and this method was published on their website (Qsrinternational.com, 2016)

Figure 3-7 below shows how Microsoft Excel was used to develop themes and sub-themes. After tweets had been labelled in NVivo, a number of nodes were exported as a list and these were imported into Microsoft Excel. These nodes were then carefully grouped in a number of themes, and sub-themes for swine flu and Ebola. However, the nodes within each of the themes initially selected were rearranged as the project progressed, for which Microsoft Excel was useful because it supported this rearrangement. When each of the thematic qualitative reports were written, each tweet was carefully anonymised to protect the identity of users.

Figure 3-7 Use of NVivo for coding tweets

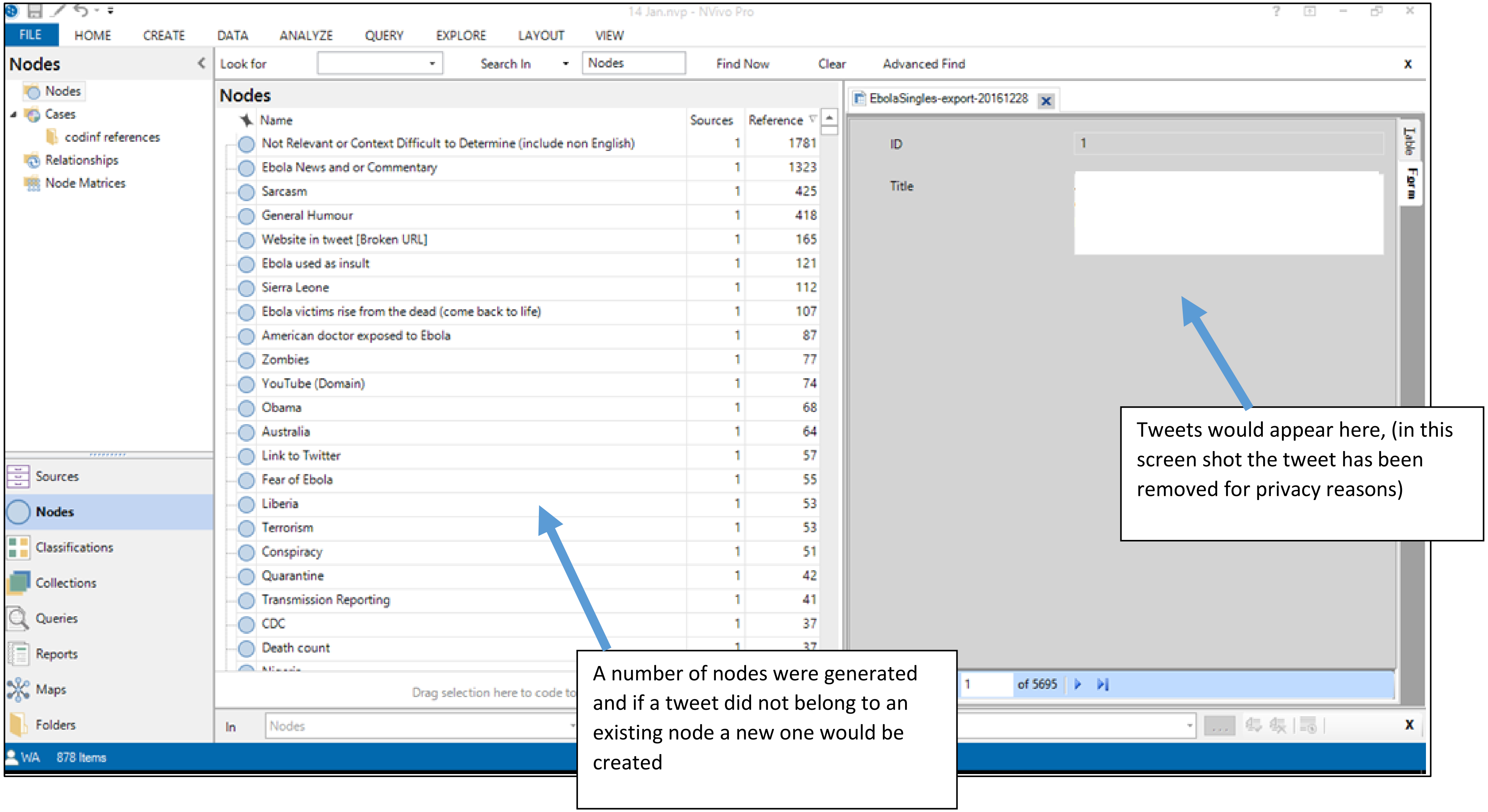
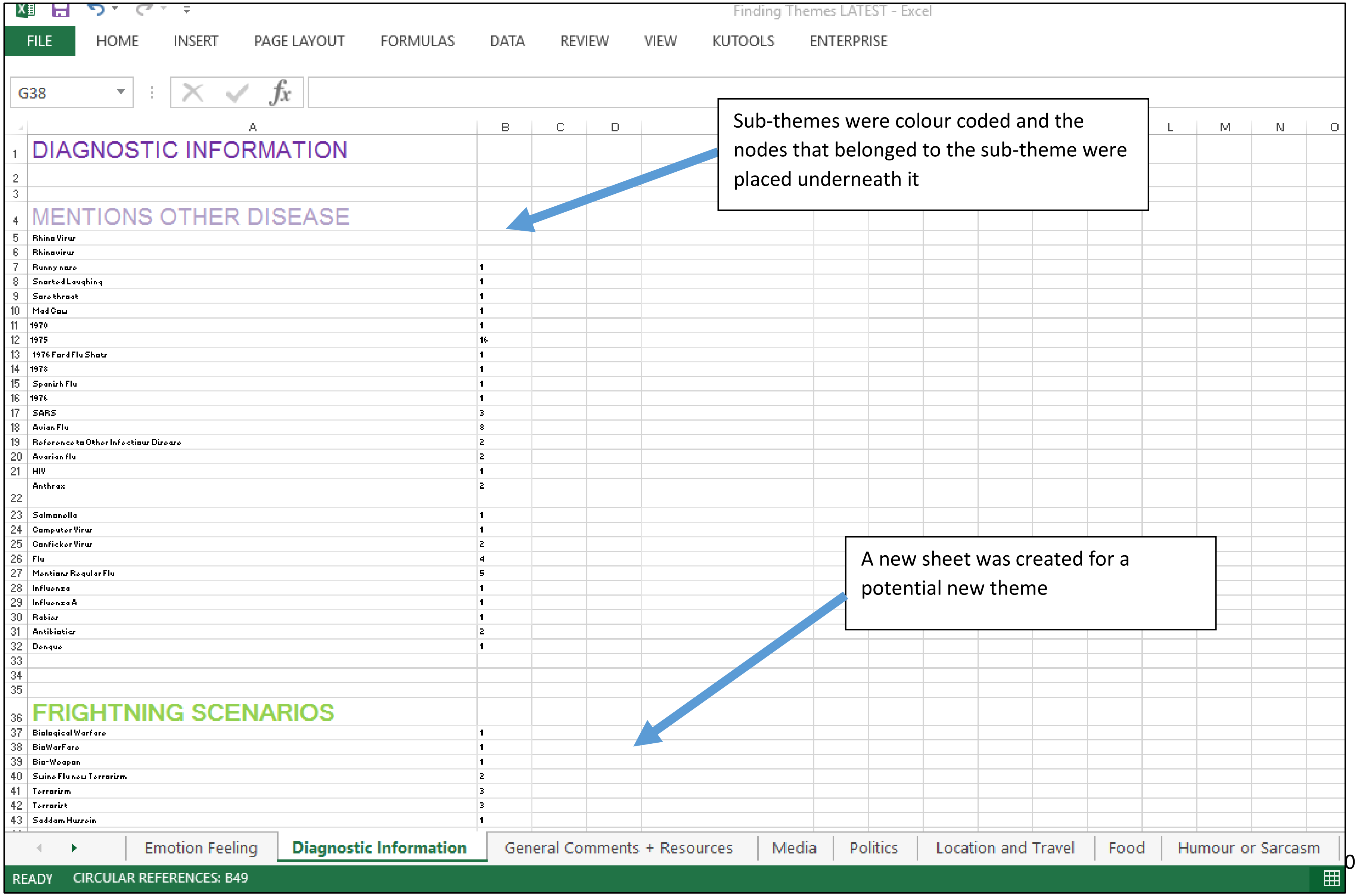


Figure 3-8 Example of Microsoft Excel use for finding themes and sub-themes from the data



3.16 Validity and Reliability

3.16.1 Intercoder Reliability

For this study, a total of 300 tweets were coded by an academic who was independent to the research project. According to Neuendorf (2002), it is not necessary to code more than 300 units when generating intercoder reliability as the statistics would be unlikely to change. In NVivo (2012), it is possible to run a coding comparison which allows another researcher to code a subset of tweets and to generate an intercoder reliability percentage agreement, as well as a statistic. The coder, a PhD student with experience in NVivo, was sourced and provided verbal instructions in how to perform the coding and the coder was able to ask questions about the nodes that were generated. However, once coding began, no communication took place between the coder and researcher. The scores for the intercoder reliability can be found in Chapter 4 Swine Flu (Section 4.13) and Chapter 5 Ebola (Section 5.9.2), respectively. It must be noted that, as this study is qualitative in nature, and tweets could have been openly coded, intercoder-reliability may not be the best reliability measure for this study. Therefore, test-retest reliability was also conducted as described in the next section.

3.16.2 Test-retest reliability

Test-retest reliability is traditionally applied in survey research and measures the consistency of respondents by asking them to complete a survey at different time intervals (Litwin and Mark, 1995). This test was performed in this study, and after having coded data on Ebola and swine flu a sample of 300 tweets were recoded using NVivo's (2012) coding comparison feature. Test-retest reliability statistics and percentage agreements were calculated and are reported in and Chapter 4 Swine Flu Case Study (Section 4.12.1) and Chapter 5 Ebola Case Study (Section 5.9.1), respectively.

3.17 Comparison of Cases

The different case studies on swine flu, Zika, and Ebola were analysed individually and then brought together for comparison in order to highlight broader similarities, and differences. The comparative method has been well established in social science research and a strength of examining individual and independent cases in qualitative research is the potential for these cases to be compared to one another. Ragin (2014), for instance, has developed and formalised a technique of holistic qualitative comparison which has been utilised for a quarter of a century, and other researchers have highlighted the benefits of cross-case analysis in qualitative research (Eisenhardt, 1989).

However, it appeared that, to the best of the authors knowledge, these methods had not previously been applied to the domains of research conducted on Twitter and specifically where the individual unit of analysis was a tweet. Previous research had applied methods such as cross-case analysis predominantly in the context of interview data (Miles, Huberman, and Saldana, 2013). Moreover, the goal of the comparison in this study was to highlight potential similarities and differences to how Twitter users responded during the outbreak period rather than develop a theory which tends to be an aim of certain cross-case analysis methods (Eisenhardt, 1989). However, the comparison in this study still contains elements of a traditional social science comparative method because each of the cases were analysed individually and then brought together for comparison purposes. This comparison can be found in Chapter 6 Discussion section 7.3.

3.18 Summary

This chapter has provided an overview for the methodological considerations of the study. This study is based on the pragmatic research paradigm; however, elements of the study were identified to be related to interpretivism (as described in section 3.2). The research methodology was primarily qualitative, and involved some quantitative elements such as counting numbers for themes. The research method was exploratory and descriptive and Twitter data was used as the primary source of data, and the chapter provided an outline of how the data were retrieved, filtered, and analysed. The next chapter details results from the analysis of tweets on the swine flu outbreak from 2009.

Chapter 4 Swine Flu Case Study

4.1 Introduction

The previous chapter outlined the methodological consideration of this study. This chapter details the results of Twitter data analyses on the swine flu pandemic corresponding to the 28th and 29th April 2009. This chapter first provides some background information on 2009 swine flu outbreak (section 4.2), then it reports on the eight prominent themes identified from the data based on an in-depth qualitative analysis (section 4.3). Throughout, the results presented in this chapter, and where applicable, the Health Belief Model was used to understand the rationale and behaviour of Twitter users.

4.2 Background

The swine flu (H1N1) virus was first identified in Mexico in April 2009 and, because of this, was also known as the Mexican flu virus (NHS Choices, 2015). The virus is called swine flu because of its genetic resemblance with a known influenza ('flu') virus that can cause illness in pigs (Davis, 2015; NHS Choices, 2015). The medical terminology for the virus is A/H1N1pdm09 (WHO, 2009b). In 2009, the virus spread rapidly from country to country, as it was a new type of flu virus to which people were not yet immune (Davis, 2015; NHS Choices, 2015). The symptoms of flu, which are caused by the H1N1/09 virus, are similar to other types of flu and can include (Davis, 2015; NHS Choices, 2015):

- Sudden fever, with a temperature of 38°C or above
- Tiredness
- Aching muscles or joint pain
- Headache
- Runny or blocked nose

Most people suffering from swine flu recover within a week, even without special treatment (Davis, 2015; NHS Choices, 2015). With regard to the treatment of swine flu, antiviral medications are available to help relieve symptoms and reduce the risk of serious complications, as well as antibiotics, which help combat bacterial infections such as pneumonia. The H1N1 virus spreads in a similar way to the common cold and other flu viruses (NHS Choices, 2015). The virus can be found in the millions of tiny droplets that come out of the nose and mouth when a person coughs or sneezes, and which can spread around up to one metre (three feet) from the source (NHS Choices, 2015). These droplets hang on to surfaces where the virus can survive for up to 24 hours, and anyone who touches the surfaces can spread the virus by then touching other objects (NHS Choices, 2015). In terms of prevention, it was noted that regular flu vaccines should be able to help protect against swine flu for those members of the general public who would have a high risk of developing the disease, such as elderly people and those with weakened immune systems (NHS Choices, 2015).

4.3 Key Events During the Outbreak

It is important to understand the context of the outbreak, specifically the events that took place during the time period under study, as this has the potential to influence the discussions that were taking place on Twitter. The selection criteria for the key events to be analysed were that they had occurred slightly before and were related to the time period of the data retrieval. The tables below were re-created from NBC's official timeline of the outbreak based on an article that they published in 2009 (Timeline: 2009 swine flu outbreak, 2009), and provide an overview of the key events that took place from 28th March to 28th April 2009. Figure 4-1 is a time-series graph which provides an overview of Twitter activity during April 28th and 29th 2009.

Table 4-1 Events from 28th March 2009 to 21st April 2009

Date	28th March 2009	13th April 2009	17th April 2009	21st April 2009
Event	• Earliest known onset of swine flu in the US	• First person known to die of swine flu. • The death occurs in Mexico.	• First cases officially reported. • One case in San Diego County. • One case in Imperial County of California.	• Warnings issued in US, to doctors about a new strain.

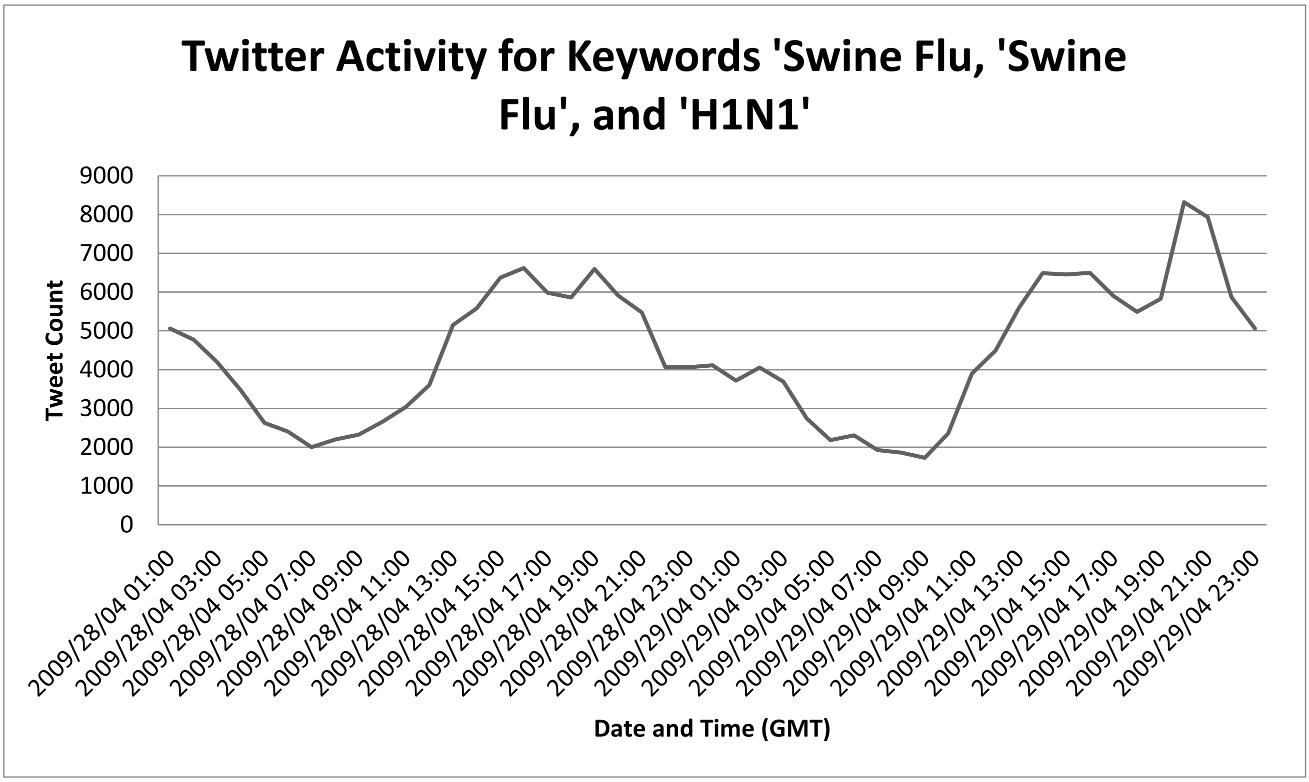
Table 4-2 Events from 24th April 2009 to 28th April 2009

Date	24th April 2009	25th April 2009	27th April 2009	28th April 2009
Event	• Eight people are confirmed to have swine flu in the US • News at this time reports that creators of Relenza and Tamiflu note that their antiviral drugs can fight swine flu.	• The World Health Organisation calls an emergency meeting. • Public Health Emergency was declared.	• The WHO raise their pandemic alert level to level 4, a move indicating that an actual pandemic is imminent.	• Confirmed cases in the US rise to 68. • Obama requests \$1.5 billion in emergency funding as an emergency measure.

Table 4-3 Events from 29th April 2009 to 30th April 2009

Date	29th April 2009	30th April 2009
Event	• The Pandemic Alert is raised to level 5 by the WHO. • Mexico notes that it has suspended non-essential services related to government and some businesses from May 1 to 5. • Egypt slaughter around 300,000 pigs during this time in order to prevent the spread of swine flu. • Child in Texas becomes first US person to die from swine flu.	• 300 Schools across the United States close in order to slow the spread of swine flu.

Figure 4-1 Time series graph displaying Twitter activity from 28/04/09 to 29/04/09.



The time-series graph above displays the tweet activity related for the data that were analysed as part of this case study. It can be seen that there was a steady stream of tweets, and a peak of tweets on the 29th of April, and this peak correlates to a number of events described in the tables above, the most significant that the WHO raised their alert level to 5, and that a child in Texas became the first person to die from the virus.

4.4 Summary of Data Selection Procedures

A summary of the data selection and justification procedures which were outlined in Methodology Chapter 3 Section 3.14 are summarised below:

- Data were retrieved from Twitter using Twitter's Firehose API and were captured using Visibrain, a licensed reseller of Twitter data.
- In total, 214,784 tweets were retrieved from the time period of April 28th and April 29th 2009 using the keywords of 'swine flu', 'swineflu', and 'H1N1' (section 3.5.1 provides an overview of the challenges of retrieving data via keywords).

- This time period was selected as Google Trends data suggested that there was a peak in related searches (Google, 2017).
- Once duplicate tweets had been removed, 102,852 single tweets remained.
- Near-duplicate tweets at a 60% threshold were removed. There were 76,783 tweets found to be single items.
- A 10% random sample of these tweets (7,679) was taken and entered into NVivo in order to conduct a systematic in-depth thematic analysis of the tweets.
- A 10% sample was taken for feasibility purposes, as it ensured that there was an equal chance of selecting each tweet and served as a representative sample of the entire dataset.

4.5 Summary of Data Analyses

The technique used to analyse the data in this study was thematic analysis (Braun and Clarke, 2006). The analysis was completed using NVivo and Microsoft Excel in line with the six phases outlined by Braun and Clarke (2006). The first stage involved becoming familiar with the data and a number of tweets were read in Microsoft Excel. The second stage involved generating a number of initial codes and this was conducted by using NVivo to highlight tweets and assign nodes to the sample of tweets. The third stage involved exporting the nodes into Microsoft Excel and searching for potential themes by examining nodes. The fourth phase of analysis consisted of reviewing themes and rearranging and discarding nodes. The fifth stage involved defining and naming the final themes that emerged from coding tweets, and the final stage involved writing the report (Section 6.7 of this chapter).

4.6 Summary of Platform Features

Ellison and boyd (2013) argued that it is important to document and highlight the features of a social media platform that is being researched. This is because features, as well as the user-interface of social media platforms, can change rapidly, and from the time of data-collection to publication platforms may have altered their features. Moreover, highlighting platform features at the time of data collection makes it possible to compare results to studies that may have conducted research when the features of a platform were different. This is also important for the purposes of this study, which was seeking to compare the onset swine flu from April 2009 with the Ebola outbreak from September 2014 as well as to the Zika outbreak from 2016. These particular years were selected as there was a peak in Web search queries (Google, 2017). Between 2009 and 2014, new features of Twitter came into place. It is therefore important to document how the Twitter platform has changed during that time, and the list below documents the changes that were implanted on Twitter after 2009 (Dredge, 2014):

- Until 2009, it was not possible to verify an account. This feature came into operation in June 2009; therefore, during the time period for which the data were retrieved to analyse swine flu tweets, there were no verified Twitter users.
- Until 2009, hashtags were not hyperlinked, so it was not possible to click on a hashtag and see other uses of it. However, it was possible to search for hashtags. The ability to search for Twitter users came into operation in July 2009.
- Until 2009, it was not possible to add Twitter users into a list, so as to see what these users were tweeting about. The Twitter list feature became operational in October 2009.
- Until 2009, it was not possible to natively retweet other Twitter users via a button. Users would share other users' tweets by adding the notation of RT followed by the Twitter user. For example, RT @User. The native re-tweet feature came into operation in November 2009.
- Promoted tweets, accounts, and Twitter placing tweets in a user's timeline as a type of advert were rolled out at different times after 2010.
- Twitter released its first mobile application in 2010. Users tweeting before this date would, therefore, be doing so on a web-browser either online or on their mobile browser.
- Twitter's user-interface was different in 2009, and it has since gone through a number of changes, e.g. the location of where tweets would be shown to users and the addition of trending topics on the Twitter homepage.
- There was no feature that could send emergency alerts to Twitter users who opted into the emergency alert service. This is a feature that came into operation in September 2013.
- Twitter has gone through a number of changes in terms of timelines and how tweets are displayed. For example, a new feature has been implemented showing tweets in chronological order under the banner 'while you were away'.

- Streaming video, animated gifs, and Twitter's partnership with Vine were rolled out at various times after 2013.
- Muting and blocking other Twitter users was different in 2009 and a number of changes have been implemented since 2013, e.g. users were able to tell whether they have been blocked by other Twitter users.

4.7 Results – Qualitative Analyses

Eight prominent themes emerged from the dataset. These are listed below and are described in Table 4-4:

- Theme A: **Emotion and Feeling**

  Tweets expressing emotion towards the swine flu outbreak. These could include fear, worry, anger, and panic.

- Theme B: **Health Related**

  Tweets discussing medical concepts such as transmission, prevention techniques, symptoms, medication and diagnosis.

- Theme C: **General Commentary & Resources**

  Tweets expressing general commentary towards swine flu.

- Theme D: **Media and Health Organisations**

  Tweets mentioning a media organisation or expressed a view towards the media.

- Theme E: **Politics**

  Tweets making reference to politics or a political figure.

- Theme F: **Country of Origin (Mexico/Travel)**

  Tweets referring to Mexico, Mexicans and travel.

- Theme G: **Food**

  Tweets referring to food such as bacon, or allude to food by mentioning kosher, halal meat. These types of food derive from pigs, which are the animals associated with the virus (Davis, 2015; NHS Choices, 2015). The vegan and vegetarian diets, which do not involve meat consumption, were also mentioned.

- Theme H: **Humour and/or Sarcasm**

Tweets displaying humour and sarcasm related to swine flu.

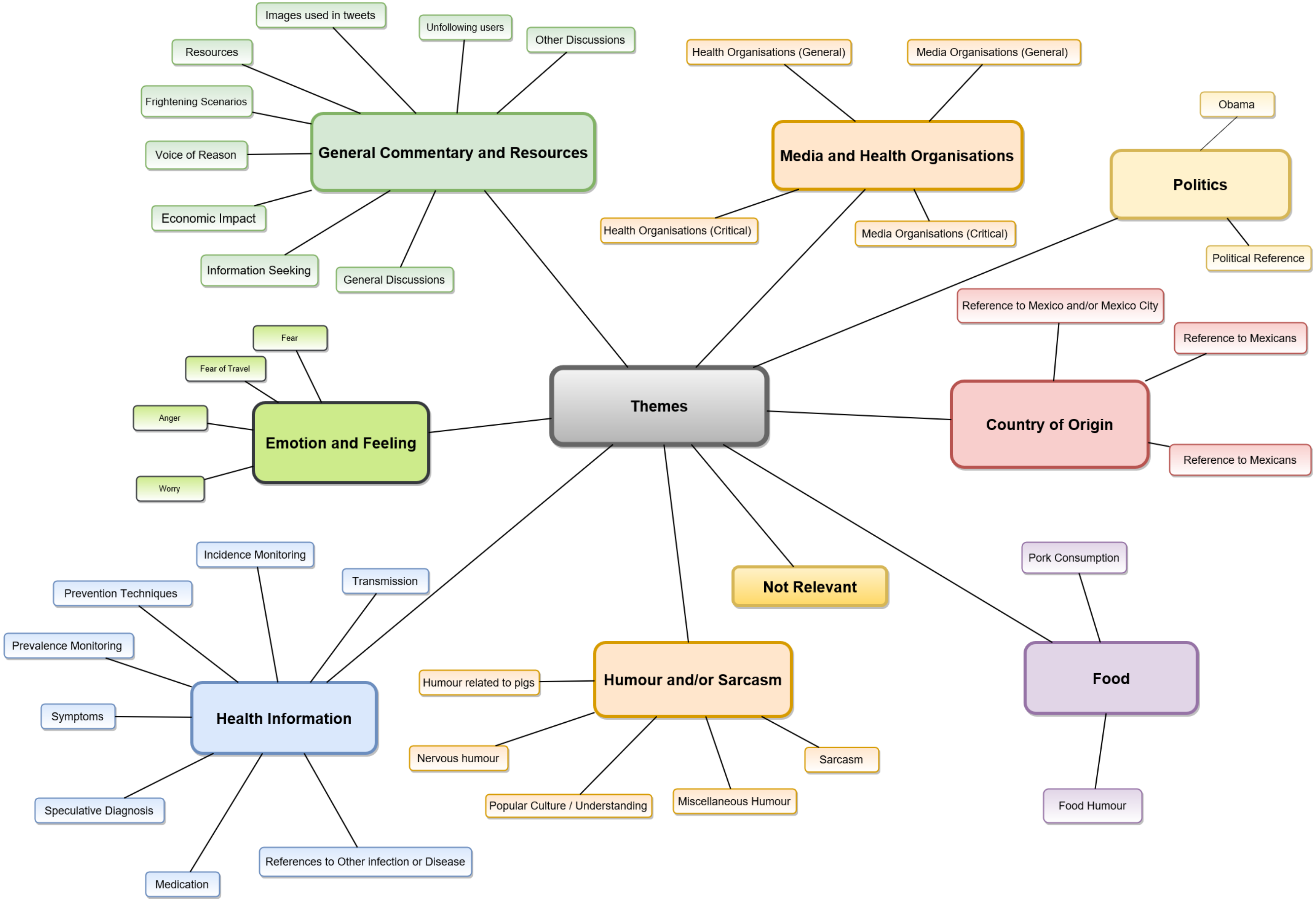
These themes, alongside the identified subthemes, are described in Table 4-4 below, and then are individually discussed in the remaining part of the chapter. In order to generate themes and sub-themes, nodes were exported from NVivo into Microsoft Excel and were then grouped together. Figure 4-2 is a diagrammatical overview of themes. A total of n=1,937 tweets were found to be not relevant and were eliminated during coding.

Although there were procedures in place to gain consent from Twitter users, it was found that it would not be possible to contact each individual Twitter user for consent due to the volume of tweets. Therefore, the tweets provided in this chapter have been re-worded in order to protect the anonymity of Twitter users. Themes were provided with labels in order to identify them in different sections. For example, Fear is labelled as *A. Fear* and the tweet illustrations belonging to this sub-theme are labelled as *A1.1, A1.2* and so forth.

Table 4-4 Overview of themes and sub-themes (swine flu).

Theme (N/%)	Sub-themes (N/%)
A. Emotion and feeling (253/4.4%)	A.1 General Fear (174/3.0%) A.2 Fear of Travel (54/0.9%) A.3 Anger (17/0.3%) A.4 Worry (8/0.1%)
B. Health Information (609/10.6%)	B.1 Transmission (22/0.4%) B.2 Prevalence Monitoring (158/2.8%) B.3 Prevention Techniques (134/2.3%) B.4 Prevention Products (126/2.2%) B.5 Symptoms (80/1.4%) B.6 Speculative Diagnosis (18/0.3%) B.7 Medication (14/0.2%) B.8 References to Other Infection or Disease (57/1.0%)
C. General commentary & Resources (2467/43.0%)	C.1 General Discussions (1826/31.8%) C.2 Information Seeking (145/2.5%) C.3 Economic Impact of Swine Flu (62/1.1%) C.4 Voice of Reason (109/1.9%) C.5 Frightening Scenarios (13/0.2%) C.6 Name Discussion (26/0.5%) C.7 Resources (42/0.7%) C.8 Images used in Tweets (36/0.6%) C.9 Unfollowing Users (2/0.03) C.10 Other Discussions (206/3.6%)
D. Media and Health Organisations (675/11.80%)	D.1 Health Organisations (general) (136/2.4%) D.2 Health Organisations (critical) (7/0.1%) D.3 Media Organisations (general) (444/7.7%) D.4 Media Organisations (critical) (88/1.5%)
E. Politics (124/2.2%)	E.1 Political Reference (81/1.4%) E.2 Obama (43/0.75%)
F. Country of Origin (Mexico/Travel) (211/3.7%)	F.1 Reference to Mexico and/or Mexico City (162/2.8%) F.2 Reference to Mexicans (43/0.8%) F.3 Reference to Borders (6/0.10%)
G. Food (428/7.5%)	G.1 Pork Consumption (336/5.9%) G.2 Food Humour (92/1.6%)
H. Humour or Sarcasm (975/ 17.0%)	H.1 Humour Related to Pigs (100/1.8%) H.2 Nervous Humour (18/0.3%) H.3 Popular Culture/Understanding (221/3.9%) H.4 Miscellaneous Humour (378/6.6%) H.5 Sarcasm (258/4.5%)

Figure 4-2 Diagrammatical overview of themes and sub-themes (swine flu)



A. Emotion and Feeling

This theme encompasses tweets which expressed emotions towards the swine flu outbreak such as: fear, fear related to travel, worry, and anger, as illustrated in the table below. At the time of the outbreak, Twitter users may have been receiving information from a number of sources, which may have had the potential to cause anxiety and fear on Twitter and among the general public.

Table 4-5 Sub-themes of emotion and feeling

Theme	Sub-themes
A. Emotion and Feeling	**A.1 General Fear**
	A.2 Fear of Travel
	A.3 Anger
	A.4 Worry

A.1 General Fear

The swine flu outbreak evoked fear in a number of Twitter users. There was a heightened sense of anxiety during this time which may have been connected to the news articles that were published and had the potential to cause fear among users. For example, users would comment on how swine flu had made them feel afraid:

'Swine Flu is scary' (A1.1)

'Heard about the Swine Flu virus and it sounds really scary' (A1.2)

'Really frightening hearing about Swine Flu' (A1.3)

'Freaking out about Swine Flu' (A1.4)

There were some Twitter users who would note that they were afraid and would post on Twitter in order to see if other users were also feeling afraid, as illustrated in the examples below:

'Swine Flu is so scary - what do other people think?' (A1.5)

'Anyone else afraid of Swine Flu?' (A1.6)

It is important to note that the age group that has been reported to use Twitter the most is 18 to 29 years, representing 36% of all users on the platform (Pew Research Center, 2016). Therefore, the comments may appear to be exaggerated, particularly to readers who are not familiar with the type of language used on social media platforms. Hence, for example, if a user was feeling afraid, it is possible that they would tweet '*I am scared to death!'*, but what the user may have meant to say is *'I am feeling afraid'*. A number of Twitter users might have been attempting to propagate fear, and a number of tweets were identified as belonging to this category due to the exaggerated language that was used in them. Examples of these tweets are illustrated below:

*'Crazy why people are not afraid of Swine Flu – it is scaring the sh*t out of me!' (A1.7)*

*'ahhhhhhhh fu**ing Swine Flu! Hope it doesn't reach me!' (A1.8)*

*'Holy sh*t man! Reports of Swine Flu everywhere! We are dead!' (A1.9)*

'Who else is scared to death of the Swine Flu outbreak?' (A1.10)

The users in the tweets above may have been attempting to perpetuate fear rather than express genuine fear. However, given the immediacy of Twitter, and the exaggerated vocabulary that was observed within the dataset of tweets, it is possible that users were experiencing a high level of fear, but were tweeting in such a way that their fear, and the actual risks, were exaggerated beyond normal proportions. This relates to a known problem in Twitter data analysis, whereby, with certain tweets, it may not be possible to ascertain the motive of the user from the tweet itself. This aspect of social media research was highlighted in a report by NatCen Social Research, which noted that users may tend to behave differently when they share information online (Beninger *et al.*, 2014). Specifically, it was mentioned that social media posts are prone to exaggerated views and impulsive comments, which may affect the validity and reliability of results drawn from social media.

A notable feature of tweets which expressed fear towards swine flu was that users wanted to know whether their fear was justified, and did this by asking how others felt. Similar to the tweet illustrated in A1.1, a user noted that they were afraid of swine flu and sought information related to how other users were feeling, as illustrated below:

'OMG, I am so scared of the Swine Flu, who else feels sick?' (A1.11)

The user in the tweet above may have been using the term sick to refer to feeling sick due to worry rather than sick with the swine flu virus itself. As noted previously, the vocabulary used in the illustration above is exaggerated. Specifically, it is a combination of Internet slang

('OMG' is an acronym for the phrase, 'Oh My God') and a request for information related to whether other users are feeling sick. There may be some tweets in which users' views should not be taken literally (for example, *A1.11* where a user notes they are scared to death, although they clear mean this figuratively), and may reflect the relatively young age group of many Twitter users, i.e. users tend to be young (Pew Research Center, 2016). However, although the tweets sent are exaggerated and may not be intended to be taken literally, users may have experienced these feelings nevertheless, e.g. fear on at least some level.

Marwick and Boyd (2014) examined Internet drama in the context of teenager conflict in the US. Whilst examining youth conflict on social media, they coined the term 'Internet drama' to refer to the idea that teenagers on social media are aware that other users can see their social media posts, and may formulate them as a type of performance, in order to appeal to other social media users. However, what is relevant to this study is that when referring to swine flu, Twitter users may be aware that their posts are being read by their followers and tweets may reflect a type of performance to their peers, i.e. their followers. This will be further explored in the discussion of results (section 7.9).

With regard to specific areas, swine flu is a disease that is not confined to a specific geographical location. The swine flu virus has the potential to reach most geographical locations on the planet due to modern commercial air travel. Swine flu is also an invisible threat, so users in affected areas may be in a continuous state of fear. Therefore, reports of swine flu may be able to incite a sense of anxiety and fear in a larger proportion of society compared to other types of news stories. Twitter users may have played on this fear and directed comments towards it, and the fears expressed on Twitter may have amplified concerns.

A.2 Fear of Travel

There were tweets which reported specific fear towards travel, such as general concern from frequent travelers and concerns that the outbreak would affect friends and family travelling to areas affected by swine flu. Additionally, Twitter users who were scheduled to be travelling to Mexico referred to how the outbreak of swine flu could potentially affect their holiday, as illustrated in the examples below:

> *'I travel a lot and I am really concerned with the Swine Flu outbreak' (A2.1)*
>
> *'I need to go on a cruise. Hope the Swine Flu doesn't come anywhere near to me' (A2.2)*

‘Swine Flu has the potential to negatively influence our Mexico City trip’ (A2.3)

Some Twitter users were unsure about future visits to Mexico, for example, a holiday or a school trip, as illustrated below:

‘Afraid that the Swine Flu will mean I am unable to travel in the holiday period’ (A2.4)

‘Hope Swine Flu does not affect our class trip aboard’ (A2.5)

In addition to tweeting about their own fear of travelling during the height of the outbreak, users also expressed fear about family members who were travelling, as illustrated in the examples below:

‘I am so scared about my mum who is going to be travelling this week’ (A2.6)

‘Leaving from Mexico early – mother was afraid I would get Swine Flu’ (A2.7)

Example A2.6 expressed fear related to the travel of a family member, which has the potential to cause stress and anxiety to other Twitter users. These types of tweets indicate that the swine flu outbreak could indeed cause fear in people who were not at risk themselves. Twitter users were unsure of how their trip to Mexico would be affected and were unsure of what steps to take, and hence they may have had an information need that was unmet, as illustrated below:

‘Trip to Mexico is next week – but now wondering what shall I do?’ (A2.8)

The user above is unsure of how their trip to Mexico would be affected due to the swine flu outbreak, and in order to satisfy this information need, they would need to read, be told and be tweeted information on potential implications of travelling to Mexico, and whether it would be advisable to cancel the trip.

There was also a Twitter user who noted that they had cancelled a planned visit to Mexico due to the swine flu outbreak:

‘Cancelled my business trip to Mexico which is next week’ (A2.9)

At the time of the outbreak, a number of flights to Mexico were cancelled, as reported in the mainstream media. This could potentially have led Twitter users to cancel planned trips to Mexico, and could explain why users were particularly concerned about travel during this time. For example, Thomson and First Choice, cancelled all outbound flights to Mexico from 28th April until 8th May 2009, and it was reported that the UK Foreign Office had issued a travel warning advising that citizens should only travel to Mexico if it was essential (Perrett, 2009). Therefore, there may have been a heightened sense of anxiety and interest surrounding travel during this time period.

A.3 Anger

There were a small number of tweets which expressed anger towards swine flu or pigs by use of vulgar language; for instance, one user directed an expletive towards swine flu, as illustrated by the example below:

*'Swine Flu can go f**k itself!' (A3.1)*

Anger from Twitter users may have derived from either the swine flu virus itself, the current outbreak situation, or the frustration about the effects the outbreak. Tweets in this sub-theme could reflect a form of anxiety based on fear, or a type of nervousness towards the swine flu outbreak. Other Twitter users noted that they were very upset after hearing about the swine flu, as illustrated below:

*'I am so p***ed off if there is a Swine Flu outbreak' (A3.2)*

The user above may have been experiencing a variety of emotions, and anger appears to have been among these. Other Twitter users may have been angry about the current outbreak of swine flu as there may have been a lot of content on Twitter during this time period, as illustrated below:

'I will punch someone in the face if I hear about Swine Flu one more time' (A3.3)

Users may have intended to direct their anger at pigs, and may have intended to be ironic at the same time. Moreover, as mentioned above, user A3.3 may also be exaggerating in order to gain attention from peers. There were few Twitter users who referred to pigs negatively, as illustrated below:

'Swine Flu is not a joke! Disgusting pigs' (A3.4)

'Best way of combatting Swine Flu – punch a pig in the face' (A3.5)

The news of swine flu may have been unsettling for users, and they may have been affected by the current outbreak situation, which could explain the motivation behind some of these tweets.

A.4 Worry

There were tweets that expressed user's worries about swine flu which stemmed from the symptoms of the virus of the outbreak situation such as the spread of the virus. Illustrations of such tweets are provided below:

'I am worried about the Swine Flu' (A4.1)

'So worried about Swine Flu! Have been washing my hands non-stop' (A4.2)

'The Swine Flu outbreak is worrying me' (A4.3)

'I have been reading up on Swine Flu – I am so worried now' (A4.4)

'I am just sitting here, and worrying about Swine Flu.' (A4.5)

'Worried about my family in Mexico'(A4.6)

During current and historical infectious disease outbreaks, health authorities have provided stress-reduction guidance and strategies. For example, during the Zika Virus disease outbreak in 2016, the US Department of Health and Human Services issued guidance on stress-management for pregnant women (Promoting Stress Management for Pregnant Women during the Zika Virus Disease Outbreak, 2017). However, evidence of such stress-reduction strategies was not observed within this sample of tweets, and in future outbreak situations of swine flu it may be beneficial to share information related to stress-reduction on Twitter. Some Twitter users had family and friends who were residing in Mexico, and their tweets expressed a hope for their wellbeing as illustrated in *A4.6* above.

B. Health Information

Tweets which discussed medical concepts such as transmission, prevalence monitoring, prevention, symptoms, medication and speculative diagnosis are branched within the health information category, as illustrated in the table below. The analysis of these tweets represent an important finding from a public health informatics perspective, as it is important to understand the beliefs citizens may have held during the swine flu outbreak. These results can feed into the development of information that would be disseminated to the general population during an infectious disease outbreak.

Table 4-6 Sub-themes of health information

Theme	Sub-themes
B. Health information	B.1 Transmission
	B.2 Prevalence Monitoring
	B.3 Prevention Techniques
	B.4 Prevention Products
	B.5 Symptoms
	B.6 Speculative Diagnosis
	B.7 Medication
	B.8 References to other Infection or Disease
	B.9 Stigma

B.1 Transmission

There were a few tweets which discussed the transmission of swine flu, such as by coughing, and human contact. Illustrative examples of these tweets include:

'Every time I see someone coughing I assume that it is the Swine Flu' (B1.1)

'I am on the coach and someone next to me is coughing #SwineFlu' (B1.2)

'Human to human contact can spread Swine Flu' (B1.3)

'Swine Flu cannot be transmitted by eating Pork' (B1.4)

Tweets in this category did not necessarily have to refer to aspects of transmission, but could also refer to scenarios in which swine flu could potentially have been transmitted, such as coming into contact with someone who had been coughing or sneezing. It appears that people were more sensitive to potential behaviours and symptoms associated with the transmission of viruses and diseases. In the context of the Health Belief Model, this could be due to the perceived severity of contracting swine flu being high, and to heightened media coverage which contributed to an increase of the perceived threat.

B.2 Prevalence Monitoring

Some Twitter users reported on the number of diagnosed cases or the location and distance between themselves and a reported case of swine flu. In this study, the author of this thesis has developed the term *citizen disease surveillance* to refer to this kind of reporting. Citizen disease surveillance is defined as the internet-based tracking, monitoring, and surveillance of a disease outbreak using tools available (paid or subscription based) on the Internet. Twitter users would use the tools available to them to monitor the outbreak and gauge the distance it was from them; illustrative examples of these tweets include:

'Just checked and Swine Flu has been reported 40 miles from north of Austin, that's getting mighty close to us!' (B2.1)

'At least 4 Swine Flu cases now in San Miguel' (B2.2)

'California has 11 diagnosed cases of Swine Flu' (B2.3)

'There might be 5 more cases of Swine Flu in New Jersey' (B2.4)

'Swine Flu just hit North Carolina' (B2.5)

'Two confirmed cases of Swine Flu reported in the UK that is not good' (B2.6)

There were many variations of the tweets above, as different geographical locations were mentioned alongside the number of reported cases. These tweets may have been sent as a type of update to the followers of the Twitter user, and could be said to be an information sharing practice. Users may have also used this practice to meet an information need by using these internet tools to monitor the outbreak, and in relation to the Health Belief Model, it may have allowed users to gauge their severity of the outbreak. That is, if swine flu was reported as being closer to the user (i.e. proximity from the Health Belief Model) then a user would be more likely to alter their behaviour. For example, in tweet *B2.1* above, the user may have had an information need (i.e. how far swine flu was from Austin), and may have engaged with a tracking tool in order to locate it.

This form of prevalence monitoring of swine flu may also play a role on the understanding that Twitter users may have when it comes to the spread of infectious diseases during outbreaks. By monitoring prevalence in this manner, users may feel safe and in control, and following news about the spread of swine flu may offer a sense of security. Moreover, Cnet, an online consumer technology news provider, published an article on the 28th April 2009 entitled *Facebook maps the swine flu hysteria* (Hoffman, 2009). The article outlined how Facebook could be used to map the discussion of the swine flu outbreak, and also provided a link to a number of other sources that could be used to track the outbreak of the virus.

B.3 Prevention Techniques

A number of users were sharing information related to prevention techniques. There were many tweets related to hand washing and the use of water and soap in order to prevent swine flu transmission, as illustrated in the examples below:

'Most effective way to protect from Swine Flu – wash your hands people' (B3.1)

'Water and soap are the best ways of preventing Swine Flu' (B3.2)

'Come on people! Wash your hands #swineFlu!' (B3.3)

'I am obsessed lately with Swine Flu – I am washing my hands a lot' (B3.4)

There were also tweets that were worded slightly differently, for example, one user noted that poor personal hygiene would spread swine flu, as illustrated below:

'Best way of spreading Swine Flu is by poor personal hygiene' (B3.5)

The Twitter user above was most likely saying the converse of what they meant, i.e. that the best way to avoid spreading swine flu is by practicing good personal hygiene. Other users provided information related to the importance of covering mouths when sneezing and coughing, as illustrated below:

'Please let's all practice some basic manner and cover mouths when coughing and sneezing' (B3.6)

Citizens who did not obey or follow the prevention guidelines illustrated above attracted ridicule and scorn, and Twitter users would come across as intolerant. For example, a Twitter user referred to a person who had not covered their mouths when coughing in negative terms, as illustrated below:

*'I would wish this stupid person on the bus would cover their f***ing mouth when coughing and sneezing!' (B3.7)*

Example B3.6 suggests that users may feel that everyone should follow the prevention guidelines issued by health authorities. Tweets related to prevention could reflect on how citizens feel other members of society should behave. People who were seen to be going against the perceived norm, as highlighted in B3.7, were singled out by others on Twitter, although they were not personally identified.

The Health Belief Model may help understand the behaviour of individuals in these circumstances. According to this model, if Person A coughs in front of Person B, then Person B's perceived susceptibility to contracting swine flu may momentarily increase, thus making Person B feel uneasy. This may explain, for example, why users refer to those people who coughed or sneezed in such negative terms. There was also a number of Twitter users who made reference to hand shaking, and avoiding hand contact, as illustrated below:

'I am not shaking anybody's hands now! #SwineFlu' (B3.8)

'Due to Swine Flu I am scared of shaking people's hands' (B3.9)

Although, it is possible that the above tweets may not be genuine and may have been intended to be ironic, they were also coded in this theme as they may provide insights into the health beliefs of Twitter users, i.e. that hand contact can transmit swine flu. The possibility of irony highlights one of the challenges of analysing Twitter data, as it is difficult to take tweets at face value because they could potentially have a number of interpretations. In the context of the Health Belief Model, not shaking hands with other people could be explained by the perceived benefit of taking such an action; that is, not shaking hands can reduce the risk of catching and developing swine flu. A number of diseases can be transmitted via hand contact, and yet do not cause users to change their behaviour. However, as the severity of swine flu is high, then Twitter users are more likely to engage in behaviours which may reduce its contraction. The Health Belief Model is useful in this context, because if the perceived severity of a disease is high (as it is in the case of swine flu) citizens are more likely to avoid certain behaviours (e.g. shaking hands). Additionally, Twitter users may have a heightened sense of perceived severity because of the geographical spread of swine flu and/or reports in the mainstream media. This could lead Twitter users to adopt certain behaviours (washing hands), and avoid others such as (shaking hands).

Other Twitter users also referred to how people had altered the way they were greeting, for example, one user noted that they had observed people to have ceased the 'normal' kiss and handshake greeting since swine flu had begun, as illustrated below:

'Just noticed that people have ceased the normal kiss and handshake greeting, since Swine Flu' (B3.10)

The user tweeting above did not disclose their location, and it may be that they were tweeting from a region where the kiss and handshake greeting was commonplace. They may have observed a reduction in the kiss or handshake greeting, had the general thought that people had ceased to kiss when greeting, or had an experience of this. To the best of the researcher's knowledge, there is no evidence to suggest that official guidance advised against a kiss or handshake greeting.

B.4 Prevention Products

There were tweets that mentioned commercial products that could be used to help prevent swine flu, such as hand sanitisers and disinfectant sprays. One user noted that they had stocked up on hand sanitiser and a lot of disinfectant spray:

'Stocked up on hand sanitizer and a lot of disinfectant spray' (B4.1)

Other tweets referred to products sold by online retailers such as swine flu prevention kits. For instance, one person suggested that users take a look at Amazon's swine flu kits:

'Check out Amazon's Swine flu kits' (B4.2)

There were other Twitter users who referred to websites selling herbal medication, and provided a link to purchase such products, as illustrated in the example below:

'Protect yourself from Swine Flu, and purchase this herbal medication [URL]' (B4.3)

The user tweeting above may have had a vested interest in selling the herbal medication: they did not provide many details about the product and only tweeted the link. Additionally, some online retailers may have noticed the increased attention to products related to swine flu prevention, and marketed these products on Twitter.

There was also a number of tweets which referenced facemasks in their tweets, as illustrated below:

'Saw some people today wearing face masks #swine flu' (B4.4)

'Waiting at the airport – afraid of catching the Swine Flu – can see people wearing masks' (B4.5)

'Hope to see more people wearing face masks today' (B4.6)

There were some Twitter users who doubted whether facemasks were effective in protecting against swine flu, as illustrated below:

'Masks are not effective in protecting from swine Flu!' (B4.7)

'Do face masks really protect you from Swine Flu?' (B4.8)

There was also sarcasm and humour around facemasks, as illustrated below:

'Swine Flu is here, stock up on the face masks people!' (B4.9)

'Whip out the face masks people! #Swine Flu' (B4.10)

On 29th April 2009, an article was published by the UK broadsheet newspaper The Daily Telegraph, which questioned the effectiveness of facemasks for protecting from the swine flu outbreak (Telegraph, 2009). There was disagreement between governments at the time as to the effectiveness of facemasks. Mexico advised its citizens to wear facemasks, and distributed them, whereas the US argued that there was no scientific evidence to suggest facemasks could protect against swine flu and did not distribute them. The article also noted that the use of facemasks would even be harmful as it could give people a false sense of security, which might lead them to neglect to wash their hands or to take risks such as venturing into crowds. A study into surgical masks and N95 respirators argued that they do have the potential to be useful, however, their effectiveness is limited by factors such as lack of training and incorrect usage (Seale *et al.*, 2009).

In examining the tweets in this section, it appeared that Twitter users held the belief that facemasks were not effective in protecting against swine flu. However, those who did see citizens wearing facemasks associated this with the swine flu outbreak, even though there could have been other reasons for people to be wearing the masks, such as air pollution. According to the Health Belief Model, if people believe that engaging in a new type of behaviour is not likely to reduce their susceptibility a disease, then they are less likely to engage in the behaviour. In this case, it appeared that Twitter users believed that facemasks were not effective in protecting from Swine so, therefore the Health Belief Model would predict that they would be less likely to wear facemasks.

B.5 Symptoms

A number of tweets referred to possible symptoms of swine flu, such as headaches. For instance, one user noted that they had a headache and thought that this may have been swine flu, as illustrated below:

'I have a headache: swine flu or normal flu?' (B5.1)

Other tweets expressed annoyance at the similarity between symptoms of swine flu and normal flu:

'It is really troublesome to know that Swine Flu symptoms are similar to regular Flu' (B5.2)

Other tweets referred to friends or family members who were experiencing swine flu symptoms:

'My dad is puking up everywhere – maybe Swine Flu' (B5.3)

There were also tweets from those who were feeling unwell and who thought they may have caught swine flu, as illustrated below:

'I am coming down with something- hope it is not Swine Flu' (B5.4)

'Think I might have Swine Flu' (B5.5)

'Seriously, I may have come down with Swine Flu' (B5.6)

Tweets regarding symptoms might reflect anxious hypochondria, as people with symptoms such as headache or cough thought that they may have caught swine flu and were speculatively self-diagnosing, as in *B5.3* above. The finding that users had difficulty differentiating between swine flu and other types of flu could feed into future public health information campaigns in response to outbreaks of infectious diseases (e.g. information such as infographics or short videos), which highlight how users would be able to differentiate between symptoms of swine flu and regular flu.

B.6 Diagnosis

Some tweets referred to diagnosed cases of swine flu such as work colleagues, schoolmates or friends/family. For instance, one user noted that they had heard news that someone in their school had swine flu and, in this case, the user referred to the person in very negative terms, as illustrated below:

'Found out someone from my school has Swine Flu, lol, isn't that disgusting. She is the only fat one in my class' (B6.1)

The tweet above may also reveal information about the individual who is tweeting, i.e. that s/he is ignorant and lacks due diligence towards other people. This tweet uses the term 'lol', a popular slang term standing for 'laugh(ing) out loud', which suggests that the user found the situation humorous. This suggests that there was stigma attached to those contracting swine flu, as well as to obesity.

There were also tweets referring more generally to diagnosed infections, as illustrated below:

'My office confirmed a diagnosed case of the Swine Flu' (B6.2)

The illustration above refers to a diagnosed case in an office setting but similar tweets were also tweeted which referred to hospitals, schools and family members being in close contact with a diagnosed case:

'If I get another patient at the hospital that says they have Swine Flu – I am going to go mad!'(B6.3)

'Son's teacher just called – received a message saying a child in her class tested positive for Swine Flu' (B6.4)

'Brother found out there is a diagnosed case of Swine Flu near his house- he is so scared!' (B6.5)

The tweets above refer to self-diagnosed cases of swine flu across places of work and study, and residential areas. It appears that people tweeting during this time experienced increased fear of swine flu when a case was reported close to themselves or their relatives. This is consistent with the Health Belief Model's prediction that perceived susceptibility to a disease increases with reported proximity. Moreover, as swine flu is perceived to be a severe disease which people would be afraid of contracting, they may feel uneasy and afraid when the virus is reported to be near them.

B.7 Medication

There were Twitter users who mentioned a number of different types of medications within their tweets, as illustrated in the examples below:

'Tamiflu can protect against Swine Flu' (B7.1)

'Use herbs to protect yourself against Swine Flu '(B7.2)

'Need to figure out how to create Relenza or TamiFlu' (B7.3)

'Homeopathy treatment was very successful against the Spanish Flu' (B7.4)

Some Twitter users suggested that alcohol was effective in protecting against swine flu:

'Best cure for Swine Flu: vodka, was in Mexico sipping vodka all of the time, and I'm fine' (B7.5)

'Rumours circulating that if you drink a lot of alcohol and sleep, you can avoid the Swine Flu' (B7.6)

As noted previously, tweets may be susceptible of multiple interpretations and it can be difficult to judge which is correct, particularly if, as is often the case, the context is unknown, for example, in tweet *B7.6* the user may have been conveying a serious point. Additionally, it can be assumed that tweets that were intended by the writer to be humorous would sometimes be taken at face value and believed.

B.8 Reference to Other Infection or Disease

There were tweets that referred to previous infectious disease outbreaks and other diseases or infections such as: the Spanish flu, SARS, mad cow, anthrax, avian flu, rabies, influenza, HIV; and also to dates of previous infectious disease outbreaks, such as 1975, 1976, and 1978 (these dates relate to the 1970's where there were outbreaks of swine flu and an overall heighted risk of an outbreak). For instance, one user referred to previous outbreaks of bird flu, and bovine spongiform encephalopathy (BSE) (also known as mad cow disease):

'So we have had the Bird Flu, the Mad Cow, and now Swine Flu!' (B8.1)

This tweet may have been written to transmit nervous humour. Other Twitter users made reference to previous diseases and were curious to know what the future would hold:

'Omg, first SARS, and now Swine Flu – wonder what is next!' (B8.2)

As mentioned in A.1, Twitter users may have been attempting to propagate fear, and the tweet above is a further example of this. Other Twitter users drew parallels with previous infectious disease outbreaks, as illustrated below:

'Wow at Swine Flu, does it remind anyone else of SARS?' (B8.3)

This type of comparison to previous disease outbreaks may influence the perceived severity of a current outbreak which could be similar/equivalent to previous infectious outbreaks, as illustrated below:

'Is Swine Flu this year's SARS?' (B8.4)

In the context of The Health Belief Model, Twitter users would be more likely to engage in certain behaviours and avoid others if the perceived severity and their susceptibility to it was high. By referring to other deadly diseases, users believed that swine flu was as harmful as previous infectious disease outbreaks, and may have raised the perceived severity of swine flu.

Some Twitter users referenced a time period, such as '1976', and the date recalled by Twitter users varied, with some users referencing incorrect dates, as illustrated below:

'Wasn't there an outbreak of Swine Flu in 1978?' (B8.5)

In fact, the swine flu outbreak was at its peak in 1976, as other users correctly recalled:

'Wasn't there an outbreak of the Swine Flu back in 1976 – thought it turned out to be a bit of an over-reaction' (B8.6)

Another user suggested that the US government over-reacted to the 1976 outbreak at the time, and that swine flu had not been as dangerous as it was made out to be:

'Thought there was an outbreak in 1976, and the government panicked at the time, and nothing happened then' (B8.7)

A Twitter user referred to a government vaccine distributed in 1975 and how it might have caused cases of paralysis:

'The government vaccine back in 1975 caused paralysis in some people, let us hope this doesn't happen again!' (B8.8)

Another user noted that in 1975 they had received a government vaccine, and that this had led them to develop allergies to certain food products:

'In 1975 it was those Swine Flu shots which activated my allergy to eggs' (B8.9)

For tweets within this sub-theme, it was interesting to see how users shaped their understanding of the current swine flu outbreak by referring back to previous outbreaks of swine flu, and other infectious diseases.

C. General Commentary

Tweets which expressed general commentary about swine flu were placed in this theme, and further divided into ten sub-themes, as shown in Table 4-7 below.

Table 4-7 Sub-themes of general commentary

Theme	Sub-themes
C.General commentary	**C.1 General discussions**
	C.2 Information seeking
	C.3 Economic impact of swine flu
	C.4 Voice of reason
	C.5 Frightening scenarios
	C.6 Name discussion
	C.7 Resources mentioned or referred to in tweets
	C.8 Images used in tweets
	C.9 Unfollowing users due to swine flu tweets
	C.10 Other

C.1 General Discussions

In line with previous research (Chew and Eysenbach, 2010), and as found in the pilot study (Chapter 5) of this project (Ahmed & Bath, 2017), there was a number of tweets that made very general comments and mentioned swine flu, and used the swine flu hashtag. Tweets in this category were quite varied, and included some of the following: tweets where the users would mention a conversation that they had overhead; tweets with the '@' notation which indicates a conversation between users; and more mundane tweets such as a user tweeting that their 500th tweet was on the topic of swine flu. Illustrative examples of these tweets are provided below:

'The guy on the bus talked about Swine Flu' (C1.1)

'@ maybe because of the Swine Flu' (C1.2)

'Hurray! My 500th tweet is on Swine Flu!' (C1.3)

'Say hello to John – was in California helping with Swine Flu relief' (C1.4)

'Interesting insight into Swine Flu today' (C1.5)

Many tweets within this category fell into C.10 Other, as they could not be categorised more specifically were quite common within this sub-theme. The Twitter user in C1.1 noted that their 500th tweet was on swine flu, which may suggest that their attitude towards swine flu was not to take it seriously. As tweets reported in this study have been anonymised, the user handle in C1.3 has been omitted. Tweets were difficult to code when they were sent in reply to other Twitter users as two or more users were in the process of discussing a topic. For example, a Twitter user may have sent a tweet related to swine flu and another user may have replied to this. In the dataset, it was not possible to see the original tweet that another Twitter user was replying to.

C.2 Information Seeking

Some Twitter users who sought information about topics such as: symptoms, medication, prevention, food and more generally whether people should be concerned with swine flu and what would happen if there was a global pandemic of the virus, as illustrated in the examples below:

'How do you know if you have Swine Flu?' (C2.1)

'Why has our Australian government stockpiled so much Tamiflu?' (C2.2)

'What are the symptoms of Swine Flu?' (C2.3)

'Can we use Lysol to fight Swine Flu?' (C2.4)

'Does Swine Flu mean that we can't eat bacon?' (C2.5)

'What can we do to stop our children from getting Swine Flu?'(C2.6)

Tweets in this sub-theme provide evidence that people were using Twitter to seek information. Analysis of this kind of Twitter data could be used by health authorities and stakeholders with an interest in health to develop material to be distributed online. Rather than monitor Twitter and address each question posted by users individually, organisations such as the WHO and CDC, could develop tweets that respond to and address users' information needs *en masse*. This method of disseminating information should be less labour-intensive than monitoring Twitter and replying to users individually.

Other tweets sought information on what would happen if swine flu were to reach pandemic levels:

'What is going to happen if Swine Flu starts to turn into a fully blown global pandemic?' (C2.7)

There were also more general tweets that sought information from Twitter in relation to whether people should worry or be afraid of swine flu:

'Do I need to worry about Swine Flu?' (C2.8)

C.3 Economic Impact of Swine Flu

A number of tweets mentioned the economic impact that swine flu was having on the jump or decline of the dollar or the stock market. Some users questioned whether swine flu was invented to distract people from the bad economic news that were circulating at the time (swine flu occurred during the 2009 financial crisis):

'Dollar goes up because of all of the Swine Flu worry' (C3.1)

'Stock market is not looking good due to Swine Flu' (C3.2)

'Swine is now in the UK, and now we think, wasn't the recession enough' (C3.3)

'Wonder if all of this Swine Flu stuff is here to try and distract us from recession' (C3.4)

At the time of the swine flu outbreak (April 2009), a global recession was taking place, and some tweets made reference to this crisis. This sub-theme suggests that Twitter users were linking the swine flu outbreak to other current events perceived as negative.

C.4 Voice of Reason

There were tweets across the dataset which acted as voices of reason and downplayed the significance of swine flu and asked fellow Twitter users to remain calm. Illustrative examples of these tweets:

'Everyone needs to calm down about this Swine Flu' (C4.1)

'The regular Flu by itself kills so many people – people need to relax #swineFlu' (C4.2)

'Receiving so many messages about Swine Flu – people really need to relax' (C4.3)

'Come on people! Get over this Swine Flu hype – not that big of a deal, and it will never be' (C4.4)

A number of Twitter users compared the risk of developing swine flu to other more likely events such as getting hit by lighting, and referred to the small number of deaths that there had been compared to the global population. Illustrations of these tweets are:

'There is higher risk of being hit by lightning than actually getting the Swine Flu' (C4.5)

'We have a better chance of catching and developing breast cancer then Swine Flu' (C4.6)

'Come on people – there have only been 100 deaths out of potential 6 billion' (C4.7)

'Not sure why everyone is freaked out, nobody in the U.S. has died from Swine Flu' (C4.8)

'For everyone that is afraid of Swine Flu – please look up the number of deaths from the regular Flu' (C4.9)

These results are interesting and surprising as, at the time, CNN wrote a news article entitled '*Swine flu creates controversy on Twitter*' (Sutter, 2009). However, the tweets above indicate that some users were actively attempting to avoid controversy by asking people to stay calm and use logic, highlighting the very small number of deaths caused by swine flu. Although Twitter users were asking for reason and logic, this does not necessarily mean that there was no controversy on Twitter, or even that these 'voices of reason' were not engaging in controversy. CNN quoted the Centers for Disease Control (CDC) as saying that people talking about swine flu on Twitter were helpful as this raised awareness among the general public (Sutter, 2009). It is possible that Twitter users who were tweeting at this time were over-loaded with information about swine flu, and this finding may be of potential interest to health authorities and stakeholders who tweet during infectious disease outbreaks. It must also be noted that the choice of title for this sub-theme does not suggest that a more muted response to swine flu was the reasonable position. This is because, if the outbreak had turned out to be much worse, it may not have been reasonable to downplay it.

A study conducted on UK participants sought to examine the UK public view and the reaction of the media and government reaction to the 2009 swine flu pandemic (Hilton and Smith, 2010). A purposive sample of 73 participants took part in focus group discussions. The study found that there was limited evidence suggesting that the public had over-reacted during the swine flu pandemic. It was found that people felt helpless in protecting themselves and that contracting the virus would be unavoidable. The study found that participants were more concerned with the vaccine rather than the virus itself.

C.5 Frightening Scenarios

Some Twitter users posed frightening scenarios in their tweets, such as "biological warfare", "bio warfare", and "biological weapons" that might harness swine flu to cause harm to the general public. There were also tweets which alluded to the idea that the swine flu infection was created by humans. Illustrative examples of these tweets are provided below:

'Swine Flu has been created by man, and released in Mexico' (C5.1)

'Rumour has it that Swine Flu is a type of beta test for bio warfare' (C5.2)

'Swine Flu could be a terrorist plot to take over the world' (C5.3)

'Swine Flu may be a plot undertaken by a sleeper cell of Al Qaeda in Mexico' (C5.4)

These tweets could represent intentional scaremongering and sensationalising of the outbreak situation. References to worst-case scenarios may also have been intended humorously.

C.6 Name Discussion

There were tweets which discussed the term 'swine flu' itself, with some users finding the name humorous, some suggesting that health authorities should have come up with a better name, and others debating whether the name was offensive, particularly the Israeli name for the disease, '*Mexican flu'*. Israel's reasons for renaming the disease related to Jewish attitudes towards pork, the consumption of which is forbidden within Judaism (Rosenblum, 2010). Since the outbreak had originated in Mexico, the disease was dubbed 'Mexican flu'. These results were surprising, because previous studies into the swine flu outbreak of 2009 had not made these observations (i.e. that Twitter users were discussing the name of the virus), an omission which could be due to the data analysis techniques employed in these studies. Illustrative examples of such tweets include:

'I am bored by Swine Flu – but at least the name is funny' (C6.1)

'I wonder if Swine Flu would get a bit more media coverage if it was called the 'extreme bacon Flu' (C6.2)

'Israel has renamed Swine Flu the Mexican Flu to ensure it does not offend them' (C6.3)

'Israelis might be racists – Swine Flu is not Kosher so they went with Mexican Flu – go figure!' (C6.4)

'Israel renames outbreak to 'Mexican Flu' (C6.5)

'Seriously, people need to get over themselves if getting offended by the term Swine Flu!'(C6.6)

'If something is named Swine Flu – how can I take it seriously?' (C6.7)

It appears that the informal name *swine flu* given to the A/H1N1pdm09 influenza outbreak of 2009 had the potential to cause confusion among Twitter users, and was a topic of discussion in itself throughout the corpus of tweets. The WHO, alongside the United States Secretary of Agriculture, reminded citizens that pork products were safe to consume during this time by issuing a statement which noted the following:

> "In the ongoing spread of influenza A(H1N1), concerns about the possibility of this virus being found in pigs and the safety of pork and pork products have been raised. Influenza viruses are not known to be transmissible to people through eating processed pork or other food products derived from pigs." (WHO, 2009b, para 1 and 2).

However, the statement was issued in May, and so the users tweeting prior to this may not have known the official stance of the WHO. Despite the clarification from the WHO, Egypt's agriculture minister ordered immediate slaughter of all 300,000 pigs in the country (ABC, 2009). There may have been additional cases of slaughtering of pigs around the world, but these were not reported to the researcher's knowledge. An implication of this finding may be that health authorities should monitor informal usage of terms used to describe infectious disease outbreaks, and seek to swiftly clarify misunderstandings across social media platforms such as Twitter.

Additionally, an article published on the 28th April 2009 by the New York Times entitled *'The Naming of Swine Flu, a Curious Matter'* highlighted that deciding how to name the outbreak had a number of "political, economic and diplomatic overtones" (Bradsher, 2009). For example, pork producers questioned whether the name 'swine flu' was appropriate as swine flu was not a food-borne illness, and certain countries such as Israel had renamed the outbreak to Mexican flu in order to prevent their citizens from having to say the word 'swine' (Bradsher, 2009). This is another reason why pork producers would question the appropriateness of the name 'swine flu'.

C.7 Resources Mentioned or Referred to in Tweets

Twitter users referred to a number of media web resources such as Direct Gov, News Week, DC Alerts, USDA, Fox Business, Rutgers Emergency Management, AZ Central, Daily Mash, among others. Users would share information and include a link to the web resource from which the information was obtained. Illustrations of these types of tweets are:

'Swine Flu is in the United Kingdom [Direct Gov URL]' (C7.1)

‘Read this story ‘Fox Business’’ (C7.2)

‘Swine Flu is in the United Kingdom ‘PC World [URL]’’ (C7.3)

‘Interesting info on Swine Flu ‘News Week [URL]’’ (C7.4)

The range of different information sources referred to by Twitter users demonstrates how many news outlets, organisations, and government services were sharing news and information related to the swine flu outbreak.

C.8 Images Used or Referenced in Tweets

There were a range of different images included in tweets such as pigs, mask wearing, a child touching lips with a pig, flying pigs, and bacon overlaid on the Centres for Disease Control website (CDC) website. These images may have been shared for humorous purposes. Illustrative examples of such tweets are:

‘This is how Swine Flu began [Image of Child kissing a pig]’ (C8.1)

‘How eventually Swine Flu passed between pigs and humans’ (C8.2)

‘How Swine Flu began… [Image of Child kissing a pig]’ (C8.3)

The image referenced in the tweets above does not contain a date and it could have been taken days, months or even years before the swine flu outbreak. From an ethical point of view, the image may have been modified without the consent of the parents of the child, who appears to be under the age of three. This has a number of ethical implications, as the child was unable to consent to the initial picture being taken and uploaded on the Internet. There are also ethical implications of Twitter users sharing the picture without consent, as well as have downloading and modifying it without consent.

Other tweets which may have been intended to be humorous involved sharing images of bacon or of a pig that had been cooked. As illustrated below:

‘Fighting against the Swine Flu ‘[Image of bacon cooking on a grill]’ (C8.4)

‘We can use apples to protect against Swine Flu [Image of roasted pig]’ (C8.5)

There were also Twitter users who would tweet images of pigs, mask-wearing, and images of bacon:

'Go away Swine Flu – [image of two men wearing face masks]' (C8.6)

'Yummy Bacon – [Image of a slice of Bacon overlaid on CDC website]' (C8.7)

C.9 Unfollowing users due to Swine Flu tweets

It was found that some Twitter users were unfollowing other users if they shared too many tweets related to swine flu, or that they themselves had been unfollowed because they were tweeting too frequently about swine flu, as illustrated below:

'Had no choice but to unfollow that user – too much panic over Swine Flu for my liking' (C9.1)

'Keep losing followers – maybe because I keep tweeting about #SwineFlu' (C9.2)

This may be of interest to social media managers and stakeholders who disseminate information related to infectious disease outbreaks on social media platforms such as Twitter, since over-sharing information related to an infectious disease could lead users to unfollow accounts, and may be perceived negatively. This could be an area for further research, i.e. developing a study which examines the optimum rate of sharing tweets related to infectious disease outbreaks.

C.10 Other

Tweets which did not fall into any of the above themes and/or where the context was difficult to determine were categorised as *Other,* and these included messages with parts written in languages other than English. When analysing social media data, it is likely that there will be tweets that are not relevant to the project. Efforts were taken to minimise this noise in the dataset by removing duplicates and near-duplicates; however, there were still tweets which were difficult to code as new nodes or did not fit existing nodes. Illustrative examples of such tweets include:

'#Freedom #SwineFlu #NSF' (C10.1)

'During Swine Flu' (C10.2)

'The more I see Swine Flu' (C10.3)

'Thanks #SwineFlu' (C10.4)

'@ maybe Swine Flu' (C10.5)

Tweets like these which did not convey meaning or that were not written in English were coded into this node. In illustrating C*10.5* the user handle has been removed in order to preserve anonymity. A number of Twitter users mentioned swine flu when replying to other users, as illustrated in C*10.5*, and these were coded into this node. There were certain tweets which were partly in English and partly in another language and it was not possible to give examples of these tweets as they are not in English and the researcher does not have the expertise to alter the words so as to prevent identification.

D. Media and Health Organisations

This theme contains tweets that mentioned a media organisation such as the BBC, CNN, ITN, FOX News, or MSNBC and/or expressed a view towards the media. Tweets which referenced health organisations such as the Centers for Disease Control and Prevention (CDC) or the World Health Organisation (WHO) were also included in this theme. Additionally, this theme included tweets which mentioned a media organisation in a critical manner, or tweets which expressed a critical view towards the CDC or the WHO, as shown in Table 4-8.

Table 4-8 Sub-themes of media and health organisations

Theme	**Sub-themes**
D. Media and Health Organisations	**D.1 Health organisations (General)**
	D.2 Health organisations (Critical)
	D.3 Media organisations (General)
	D.4 Media organisations (Critical)

D.1 Health Organisations (General)

Twitter users made reference to health organisations such as the Centers for Disease Control and Prevention (CDC) and the World Health Organisation (WHO). Some of these tweets also included information from health organisations and some users would also provide a URL, illustrative examples of such tweets include:

'CDC says you should frequently wash your hands [URL to CDC website]' (D1.1)

'CDC say there is a vaccine for the Swine Flu outbreak' (D1.2)

'Very useful information on Swine Flu from the CDC [link to CDC]' (D1.3)

'WHO now report there have been 40 cases of Swine Flu in the U.S.' (D1.4)

'WHO were saying the outbreak of Swine Flu is not containable' (D1.5)

There were also Twitter users who referred to how the WHO had raised their alert to level 4. WHO alert levels reflect the severity of an outbreak and a higher alert level indicates a greater danger. Level 1 of a pandemic is when a virus is known to have spread in animals only, and level 4 is when an animal virus spreads from human-to-human. The different phases are summarised in Table 4-9 below:

Table 4-9 WHO alert level description

Phase	Description
Level 1	There is normally no indication that viruses in animals can cause infectious in humans.
Level 2	An influenza virus from an animal has infected humans, and is considered a pandemic threat.
Level 3	An animal or human-animal influenza virus may have caused isolated infections, but not at the level of a community level outbreak. There may be a chance that the virus can spread from human to human in restricted circumstances, e.g. hand contact.
Level 4	It has been confirmed that the transmission of an animal or human-animal influenza can cause community level outbreaks. There is a significant increase in the risk of a pandemic.
Level 5	There has been human-to-human spread of the virus in at least two countries in a single WHO region. Phase 5 will indicate that a pandemic is imminent.
Level 6	This pandemic phase is distinct from the other phases as there is a community level outbreak in at least another country in a different WHO region. This phase also indicates that a global pandemic is occurring.

Adapted from: WHO (2009a). Current WHO phase of pandemic alert for Pandemic (H1N1) 2009. *[Online] Available at: http://www.who.int/csr/disease/swineflu/phase/en/[Last Accessed 09/03/17]*

At level 4, the WHO note that there will be a significantly higher risk of a pandemic, and the aim is to contain the spread of the virus as much as possible (WHO, 2009a). The WHO alerts may have implications within the Health Belief Model, as they may have influenced how Twitter users perceived the threat of swine flu. This is because an increase in the alert level may be associated with the outbreak becoming more of a severe threat. Illustrative examples of such tweets include:

'The WHO pandemic alert is now on Level 4' (D1.6)

'Swine Flu now on Level 4 – not good news!' (D1.7)

'WHO Alert now on level 4 – that's pretty scary' (D1.8)

The news of the WHO raising the alert level caused fear among some Twitter users, possibly because of the perceived severity of the news. Other users responded with humour, such the one below referring to a fictitious CDC warning with an image of a child touching lips with a pig, as illustrated below:

'How not to catch Swine Flu – do not do this [URL to fictitious CDC URL]' (D1.9)

A finding that may be of interest to the WHO is that news of alert levels rising were shared on Twitter and caused concern among users. That is, Twitter users in tweets *D1.6* to *D1.8* claimed that the increase in alert levels caused them to feel afraid.

D.2 Health Organisations (Critical)

There were Twitter users who were critical towards the WHO, for example by suggesting that the WHO were behind in their reporting on the outbreak, that it was not updating the public enough, and that its reports of the alert level being raised to level 4 were without an explanation. Illustrative examples of such tweets include:

'The WHO are a step behind every time!' (D2.1)

'Looking at the number of deaths from Swine Flu – odd that the WHO seem so relaxed about international travel etc.' (D.2.2)

'Isn't the alert level system pointless without an explanation from the WHO? #Swine Flu' (D2.3)

There were also Twitter users who criticised the CDC for not providing enough information in relation to the location of diagnosed cases, as illustrated below:

'Why isn't the CDC telling us were the people who have been infected with Swine Flu come from?' (D2.4)

'Laughing at the CDC guidelines for Swine' (D2.5)

'*Still there is no active Surgeon General & no head of the CDC, all during the Swine Flu outbreak' (D2.6)*

'CDC tells us stay home if we are sick – yet 50% of U.S. population does not have sick days!' (D2.7)

Other Twitter users noticed a newly created Twitter account '@CDCEmergency' providing updates related to swine flu. Illustrative examples of such tweets include:

'Pretty freaked out about Swine Flu, and now we have '@CDCEmergency providing updates!' (D2.8)

'I am now following @CDCEmergency – you should also consider this to be informed' (D2.9)

'Are you afraid Swine Flu? Follow @CDCEmergency! You can track the virus and also receive tips for staying healthy!' (D2.10)

Twitter users who took a critical stance towards the WHO and CDC may have done so out of a wish to see health organisations disseminate information more proactively, keep the public up to date, and give more explanation and information when, for example, alert levels for swine flu were raised from level 3 to level 4. According to the Health Belief Model, this would have played a particularly important role in the case of swine flu because of the perceived severity of the disease. The results around Twitter users taking a critical stance towards the WHO raising alert levels may be of interest to the WHO. In a potential future outbreak of swine flu and other outbreaks, the WHO could offer more guidance on the meaning of alert levels, and better convey what it means when alert levels increase, and the potential implications this has on citizens. The tweets where users are critical towards health authorities may be of interest to the WHO and CDC; for example, in tweet *D2.4*, a user criticises the CDC for not providing enough information to the general public. In future outbreak situations perhaps the CDC could monitor Twitter for expressions of need for health information and attempt to address these needs.

D.3 Media Organisations (General)

Tweets which mentioned the mainstream media were mostly about sharing news articles. This could be described as an information sharing practice i.e. that Twitter users were keen to share information related to swine flu with their followers. Illustrative examples of such tweets are:

'You can't get Swine Flu from eating Pork [BBC News Article]' (D3.1)

'Swine Flu is in UK – [BBC News Article]' (D3.2)

'Israel renames Swine Flu as the Mexican Flu [ABC News Article]' (D3.3)

'Pharma stock go up [CNBC News Article]' (D3.4)

'Mexican government looking into Veracruz [CNN Article]' (D3.5)

It can be seen that the types of news stories that Twitter users were sharing varied, and related to different aspects of the swine flu outbreak. An article published by the CNN was particularly popular at that time (2009), and the article was shared frequently by Twitter users, as illustrated below:

'Swine Flu Creates Controversy on Twitter [CNN News Article]' (D3.6)

The article mentioned in the tweet above was shared by a number of Twitter users. One possible reason for the popularity of the article on Twitter could be that it mentioned Twitter in its title, and this may have made users of the platform more likely to engage with it. The article criticised Twitter's role during the outbreak, and noted that that users on the platform were propagating fear (Sutter, 2009).

D.4 Media Organisations (Critical)

A number of users were critical of media organisations, and felt that the mainstream media were using scare tactics, causing unnecessary panic, and using swine flu as a device to attract more views and readers to their websites. Twitter users questioned the severity of the outbreak based on the increased media coverage. There were other users who were critical of coverage from the UK press, i.e. that they were publishing fear-mongering articles (online and offline) related to swine flu. Illustrative examples of such tweets include:

'Will Swine Flu actually get as bad as what the media want us to believe?' (D4.1)

'Media has just been blowing the Swine Flu out of proportion – wonder what the next sensation will be?' (D4.2)

'Is it just a bit of media hype or an actual threat? A 100 dead out of say 6 billion' (D4.3)

'Newspapers in the UK and Ireland are pushing scary coverage related to Swine Flu' (D4.4)

'The current hysteria around Swine Flu is due to the all this hype from the media' (D4.5)

It is possible that the mainstream media published articles related to swine flu with sensationalist headlines with the sole aim of attracting more readers. Some users specifically called out news media organisations by tweeting them and asking them to reduce media coverage of swine flu, for example, a Twitter user asked whether CNN could reduce coverage on swine flu, as illustrated below:

'@CNNBrk you should reduce your coverage of swine Flu – you will create unjustified panic' (D.4.6)

Other illustrative examples are:

'Swine Flu has me terrified, I think media is making everything worse!'(D4.7)

'We must ask are we communicators introducing too much fear around Swine Flu? We should educate and that is important, but are we doing more?' (D4.8)

It appears that the Twitter user above felt that, although the media should be reporting on swine flu to educate, some organisations were using scare tactics and increasing panic among the public. These findings may be of interest to media organisations that were covering the swine flu outbreak at the time, as it appears that Twitter users were voicing concerns about increased media coverage. Moreover, the media might conceivably have undone the efforts by health authorities to inform the public by generating so much media reportage and content that users became sceptical as to whether the threat was as bad as it was made out to be.

These findings are similar to a previous study (Hilton and Smith, 2010) conducted on focus groups discussing public views of the UK media and government reaction to the swine flu. The authors found that across the three focus groups the media had been criticised for over-reporting the swine flu pandemic. The extracts from the study are particularly relevant as participants would express views such as:

> "Emma: they've [the media] set about and managed to get everyone, or the majority of people into quite a panic about the whole thing. Sarah: They do it on purpose whip up the panic and anxiety in people" (Hilton and Smith, 2010, p.6)

Another group of users within the study expressed the following views:

> "Ellie: At first they were basically saying 'that we're all doomed, 'That's the impression they were giving, and I just found that was very unnecessary. Sophie: Yes the intent to cause hysteria in people. I don't find it myself, but I've seen other people lapping it up and worrying…" (Hilton and Smith, 2010, p.6).

The present study further adds to existing evidence on how people felt that the media were blowing the swine flu pandemic out of proportion. This may have important implications because members of the general public may have brushed off media reports as overstated, which may have lowered the perceived threat of swine flu among the general public.

E. Politics

Tweets which made reference to politics were placed into the theme 'Politics' and included sub-themes of 'Political Reference', and the then US president Obama, as shown in Table 4-10.

Table 4-10 Sub-themes of politics

Theme	Sub-themes
E. Politics	**E.1.Political Reference**
	E.2 Obama

E.1 Political Reference

Some tweets would contain political references to the government and particularly to the Republican and Democratic parties in the United States, as illustrated below:

'Do the republicans want the Swine Flu to win?' (E1.1)

'We only get Swine Flu when the democrats are in power' (E1.2)

'I think you would be wasting your time if you were debating whether to blame Democrats and/or Republicans for Swine Flu' (E1.3)

'Swine Flu has jumped the border – and what has our government done? Wash your hands they say – like using a tire gauge to save planet' (E1.4)

Users who supported one particular political party and/or political figure may have used swine flu to attack the opposition party. For instance, in tweet *E1.1*, a user attacks the Republican Party, and in tweet *E1.2*, a user attacks the Democratic Party. As illustrated in *E1.3,* Twitter users were aware of the blame directed at political parties, and were critical towards users who were seen to be attaching blame to the Republican and Democratic parties.

E. 2 Obama

Twitter users mentioned political figures, and the most referred-to political figure was United States President Barack Obama, who was in office during the 2009 swine flu outbreak. Tweets on Obama varied and included humorous tweets mocking Obama, reporting on an event involving Obama, or referring to a news article related to him. Illustrative examples of such tweets include:

'Wonder if there is a correlation between Obama's PIG filled stimulus and outbreak of Swine Flu?' (E2.1)

'President Obama requests 1.5 billion dollars to fight Swine Flu [URL]' (E2.2)

'Barrack Obama did not catch Swine Flu when he visited Mexico [URL]' *(E2.3)*

'Swine Flu cases up 40 - Obama asks for calm [URL]' (E2.4)

Obama was mentioned by a number of Twitter users during the swine flu outbreak as the mainstream media were reporting statements by Obama, and these articles were tweeted by Twitter users following the outbreak. In addition to a number of news articles that mentioned Obama (illustrated above), Twitter users would also mention Obama without reference to a specific news article, and illustrative examples of such tweets are:

'Wonder how Obama will blame Bush for Swine Flu' (E2.5)

'Obama shouldn't worry about getting Swine Flu from Mexico' (E2.6)

'How does Obama respond to Swine Flu?' He would like to spend more money!' (E2.7)

These types of tweets, presumably derived from US citizens, have had an interest in the president. Obama had visited Mexico during the outbreak of swine flu, and this may explain why users were tweeting about the likelihood of Obama developing swine flu (Pilkington, McGreal and Rice, 2009). Additionally, during this time, one of Obama's security guards was diagnosed with swine flu, and the White House released a statement reassuring the public that Obama did not have swine flu (Pilkington, McGreal and Rice, 2009).

Users either wrote neutrally about Obama and made general comments, or were critical of the president and used the swine flu outbreak situation to attack him. These findings may indicate that US presidents, or other political leaders, are influential figures during infectious disease outbreaks. These findings may be of interest to health authorities, and the communications team at the White House, as Twitter users may have information needs related to the safety of the US president in future disease outbreaks.

F. Mexico

Tweets that mentioned Mexico, Mexicans, and which referenced borders were coded in this theme, and an overview of these sub-themes are provided in Table 4-11.

Table 4-11 Sub-themes of Mexico

Theme	Sub-themes
F. Mexico	F.1 Reference to Mexico and/or Mexico City
	F.2 Reference to Mexicans
	F.3 Reference to Borders
	F.4 Stigma

F.1 Reference to Mexico

A number of Twitter users referred to Mexico, where the 2009 swine flu outbreak is reported to have originated. Users blamed Mexico for the swine flu outbreak, and made jokes about Mexico:

'Just heard about Swine Flu – not cool from Mexico' (F1.1)

Swine Flu sound like something you get when on vacation to Mexico' (F1.2)

'Swine Flu first- then an Earthquake – proof that God hates Mexico' (F1.3)

'Mexico City has such bad luck!' (F1.4)

There were also news articles published at that time which referred to Mexico and Mexico City:

'Many people attempting to leaving [sic] Mexico as death toll rises [News Article]' (F1.5)

'Suspected deaths in Mexico rise to over 100 [News Article]' (F1.6)

'Swine Flu postpones X-Men film Mexico City Premiere' (F1.7)

'Number of cities in Mexico have been quarantined' (F1.8)

The swine flu virus originated in Mexico City, and this may explain why Twitter users and media outlets were mentioning Mexico and Mexico City in relation to swine flu. This sub-theme could also reflect Twitter users' interpretations of Mexico-US relations, which will be further explored in the comparison chapter. The findings might indicate that the country of origin of an infectious disease outbreak will generate interest from Twitter users, and this information could be utilised by health authorities who disseminate information during outbreaks. Additionally, certain information needs may arise from citizens related to the country of origin of an infectious disease outbreak.

F.2 Reference to Mexicans

There was a number of tweets which referred to Mexicans generally, for instance, the Mexican population, or the Mexican government, and there were others which were expressing negative views towards Mexicans, illustrative examples of such tweets include:

'All the Mexicans out there keep a look out for the Swine Flu – it's come for you' (F2.1)

'Dad called tonight and was warning me about dangerous swine Flu from dirty Mexicans' (F2.2)

'Query: A fat Mexican pig made my lunch today – wonder if I will catch Swine Flu' (F2.3)

'Keep hearing about Mexican Swine Flu? Do we stop kissing pigs or Mexicans?' (F2.4)

'May have caught Swine Flu from Mexican Pork Chops' (F2.5)

'Saw a Mexican on TV made me sneeze' (F2.6)

These tweets show that the tweeters otentially lacked knowledge of how swine flu is transmitted. The tweets also contain elements of stigma and anti-Mexican feeling. These tweets may also express amusement as *F2.3, F2.4, and F.2.6* appear to contain elements of humour. This information could potentially be used to feed into the development of health information to be disseminated on social media platforms such as Twitter. For example, tweets with images such as an infographic or a video, which might help to humanise the outbreak.

F.3 Reference to Borders

There was a number of Twitter users who referred to borders and Mexicans in negative terms during the swine flu outbreak, as illustrated below:

'Dam those Mexicans who brought pigs across the border '(F3.1)

***'**How is Mexican Swine Flu crossing border? All that money on a wall for what purpose?**'** (F3.2)*

Other Twitter users went further and requested for the border between the US and Mexico to be closed:

***'**F**king Mexicans – why don't we just close the border until Swine Flu is over?**'** (F3.3)*

'Close Mexican border because of Swine Flu' (F3.4)

Presumably, Twitter users were asking for the border between the US and Mexico to be closed as they thought that this would halt the spread of the outbreak. However, there is little evidence to suggest that border control would have helped contain the swine flu outbreak. This sub-theme has also highlighted the historic issues between US and Mexico with regard to border control, which will be further explored in the discussion of results. The points identified within this sub-theme could potentially be used in devising information which could be disseminated in case of a future disease outbreak. For example, tweets could inform users of the instances in which border control is an effective measure.

G. Food

Tweets which referred to food such as bacon or pork, or alluded to food by mentioning kosher, halal meat, vegans and vegetarians were included in the pork consumption sub-theme. These results were surprising, as previous studies into the swine flu outbreak of 2009 had not made these observations, which could be due to the data analysis technique employed in this study, i.e. thematic analysis, which may have enabled a more extensive understanding. Tweets which were humorous in nature were coded within the food humour sub-theme, as shown in Table 4-12.

Table 4-12 Sub-themes of food-related tweets

Theme	Sub-themes
G. Food	G.1 Pork Consumption
	G.2 Food Humour

G.1 Pork Consumption

A number of Twitter users referred to consuming pork or bacon, and whether it was acceptable to eat them:

'Does anyone know if it is OK to eat pork?' (G1.1)

'Can you actually catch Swine Flu from eating pork?' (G1.2)

There was also a number of users who stated that they had stopped consuming bacon and pork:

'Swine Flu is so disgusting – so glad I stopped eating Pork' (G1.3)

'I am not convinced I should be eating Bacon' (G1.4)

Other users appeared to know that it was not possible to catch swine flu from eating pork:

'Really wish that people knew you can't catch Swine Flu from eating pork' (G1.5)

'Eating pork does not transmit Swine Flu [BBC NEWS URL]' (G1.6)

The Health Belief Model predicts that Twitter users may have altered their behaviour due to personal beliefs held about swine flu. Users may have altered their behaviour and ceased to consume bacon if they thought that eating bacon would increase their risk of developing swine flu.

There were also users who referred to the dietary requirements of Muslims, Jews, vegans and vegetarians:

'Stupid Swine Flu – Maybe the Jews and Muslims were right about this one?' (G1.7)

'If you get Swine Flu – are you still adhering to Kosher?' (G1.8)

'How angry will Jews and Muslims be if they were to get Swine Flu?' (G1.9)

'Imagine how you would feel if you were Vegetarian and you got Swine Flu' (G1.10)

Users may have been sharing the tweets above for humorous purposes and to display irony. The similarity between vegans, vegetarians, Jews and Muslims is that none of these groups consumes meat that comes from pigs. However, it appears that some users may have thought that by adhering to a vegan or vegetarian diet would reduce their risk of developing swine flu. Illustrative examples of such tweets are:

'If you're a vegan, you can't catch Swine Flu' (G1.11)

'Entire world is on alert due to Swine Flu – glad I am a Vegetarian' (G1.12)

'Wish I adhered to the Vegetarian diet' (G1.13)

There is no evidence suggesting that adhering to a diet free of meat, particularly pork, would lower the risk of catching and developing swine flu. The information related to users questioning whether it was possible to catch swine flu from eating pork could be used by health authorities to disseminate information to address this information need, i.e. that it is not possible to develop swine flu from eating pork. One potential implication of these results is that when Twitter users are unsure of whether a specific food product can transmit a disease. This uncertainty should be addressed swiftly. This could be achieved by monitoring Twitter and disseminating content appropriately. Additionally, in case of future outbreaks of the swine flu disease, health authorities may wish to call for swine flu to be referred to by its medical terminology.

G.2 Food Humour

Although there are separate sub-themes of humour and sarcasm, it was found that specifically food-related humour was shared on Twitter frequently:

'Swine Flu gives my fear of pork chops justification' (G2.1)

'Swine Flu is not going to get to me – eating that bacon!' (G2.2)

'Doing my bit to halt the Swine Flu – Eating a sausage for my dinner' (G2.3)

'Swine Flu and Tacos...Yummy' (G2.4)

'Swine Flu seems like a conspiracy by the Muslims to stop us from eating bacon sarnies' (G2.5)

'For breakfast I had a pulled pork and some waffles – in your face Swine Flu!' (G2.6)

'All this talk of Swine Flu it makes me so hungry about eating ham' (G2.7)

Tweets in this sub-theme and those in *G1 Pork Consumption* suggest that some Twitter users held the belief that eating pork meat would transmit swine flu. Much of the humour in this theme is derived from satirising the belief that pork and bacon could transmit swine flu. This belief may have stemmed from the fact that the H1N1 virus was informally named '*swine flu'*. As noted in *C.6 Name discussion,* health authorities could potentially monitor social media platforms such as Twitter for informal names of infectious disease outbreaks, so that they would be in a position to offer swift clarification if an informal term was causing confusion among Twitter users.

H Humour and Sarcasm

This theme contains tweets which displayed general humour and sarcasm about swine flu, and includes sub-themes of humour specifically related to pigs, nervous humour, popular culture and understanding, sarcasm, and general humour as shown in Table 4-13.

Table 4-13 Sub-themes of humour- and/or sarcasm-related tweets

Theme	Sub-themes
G. Humour and Sarcasm	**H.1 Humour related to pigs**
	H.2 Nervous humour
	H.3 Popular Culture/Understanding
	H.4 Sarcasm
	H.5 General Humour

H.1.1 Humour related to pigs

The *A/H1N1pdm09* virus is known as swine flu as the virus resembles an influenza virus which causes illness in pigs, and not because pigs host and transmit the virus. However, because of the name 'swine flu', it appears that many users had the belief that pigs were responsible for the virus. This belief emerged in other themes and sub-themes such as *G.Food*, and *C.8 Images Used or Referenced in Tweets,* and *B.3 Prevention*. A number of tweets were centred on humour that revolved around pigs, bacon and pork meat. Illustrative examples of such tweets are:

'Trying to defeat the Swine Flu by hitting a pig right in the face – justice will be served' (H1.1)

'Wondering any connection between Swine Flu and government pigs!' (H1.2)

'Swine Flu – sounds like the revenge of the world's tastiest meat' (H1.3)

'Because of the Swine Flu infection our library is going to remove all copies of three little pigs' (H1.4)

'Wonder if I should stop eating pork – or whether you can only catch it from kissing piglets' (H1.5)

'Oink, Chew, Chew Oink – Swine Flu!' (H1.6)

'Can you get Swine Flu by looking only at pigs?' (H1.7)

There were some Twitter users who shared information related to a number of Starbucks coffee shops closing in Mexico City, and claimed that Starbucks coffee itself could spread the disease. There were no government/mainstream media reports of Starbucks closing any of their coffee chains in Mexico City, and therefore this is likely to be false information and/or an attempt at humour. Illustrative examples of such tweets include:

’10 Starbuck coffee shops close in Mexico City’ (H1.8)

‘Associated Press reported that Starbucks coffee is definitely spreading Swine Flu’ (H1.9)

One interpretation of Twitter users sharing this kind of humorous tweets is that they may have not taken the outbreak of swine flu seriously. Infectious disease outbreaks have the potential to cause a high number of fatalities in a short amount of time and are taken seriously by health authorities, who are best able to understand the severity of the outbreak. However, Twitter users on average are aged between 18 and 24 (Pew Research Centre, 2016) and may not be in a position to fully grasp the seriousness of the situation due to lack of information (i.e. they may have dormant information needs). Health authorities could target Twitter users who share this type of content and educate them on the seriousness of the outbreak.

H.2 Nervous Humour

Some tweets appeared to express or refer to nervous humour:

Keep sneezing! Might get Swine Flu and die’ (H2.1)

‘My housemate has Swine Flu – I am so dead!’ (H2.2)

‘Came home put on the news- we will all get Swine Flu and die – how exciting’ (H2.3)

‘A lot Swine Flu Jokes – wonder if this is nervous humour by users?’(H2.4)

These results could potentially be used by health authorities to disseminate information related to methods of managing stress and anxiety during an infectious disease outbreak. However, if the tweets are intended humorously, this may not be necessary.

H.3 Popular Culture/Understanding

A number of Twitter users referred to popular culture in various ways, including Hollywood films, television programs, characters from film or television, and zombies. Zombies are often depicted in Hollywood movies in connection with a virus spreading around the world. For example, in the British horror film ‘28 Days Later’, the plot centres on the rapid spread of an infectious disease outbreak where humans are transformed into flesh-eating zombies. Illustrative examples of such tweets include:

‘We are watching the situation room – character just mentioned Swine Flu!’ (H3.1)

’28 days later - #SwineFlu’ (H3.2)

‘Wondering if this Swine Flu stuff will all turn into a bit of a disaster film’ (H3.3)

Representations within popular culture, such as Hollywood films, may have been the only reference point for users hearing news of a pandemic. These representations may have played into Twitter users’ developing understanding of the outbreak. There were a number of mentions of zombies and a zombie apocalypse, as illustrated below:

‘I am so ready for when the Swine Flu starts to turn people into flesh eating zombies’ (H3.4)

‘Swine Flu will lead to a zombie epidemic’ (H3.5)

‘What is [sic] Swine Flu leads to a zombie epidemic?’ (H3.6)

‘The dead might rise and start a zombie apocalypse #swineflu’ (H3.7)

‘Got this cross-bow pre-order! [URL to checkout page] – take that Swine Flu!’ (H3.8)

Typically, in Hollywood movies or television programs which depict zombies, uninfected humans use their resources to eliminate the infected, as referenced in H3.8. This type of thinking portrayed in popular culture may have been unpleasant to read for those who were suffering with swine flu. This sub-theme also highlights that Hollywood film narratives may influence how the general public understands infectious disease outbreaks. A potential reason for this may be that, as Twitter users initially hear the term pandemic in TV and movies, this becomes their only reference point for real infectious disease outbreaks. Health authorities may be interested in these results as they demonstrate the influence popular culture has on Twitter users, and may wish to monitor popular culture for potential misinformation, and to disseminate information on Twitter that would correct this.

H. 4 Humour

There was substantial overlap between the sub-theme of humour and many of the other sub-themes outlined in this chapter. However, some tweets were solely categorised into the sub-theme of humour, as there were users who conveyed humour within their tweets and referenced the swine flu outbreak, as illustrated below:

'Oh no wearing my favourite strappy heels and feels like they have grown blades, my toes or [sic] hurting – maybe it is Swine Flu' (H4.1)

'Wonder if the air condition [sic] will keep me safe from the Swine Flu?' (H4.2)

'Had a quick scan of the news – interesting how everyone they are showing with Swine Flu is ugly!' (H4.3)

'Media has lost it over Swine Flu! But I bet those prices to Mexico are cheap now' (H4.4)

'Swine Flu is like a terrible housemate – never does the dishes' (H4.5)

'Swine Flu is like someone not billing your credit card payment – and then raising your interest rate' (H4.6)

Twitter users may feel that it is acceptable and even normal to tweet humour in relation to swine flu. Previous evidence-based research (Chew and Eysenbach 2010), as well as a pilot study that was conducted as part of this research (Ahmed and Bath, 2017), found that humour was often conveyed in tweets. A potential implication of these results is that, when a situation such as an infectious disease outbreak arises, citizens engaging with Twitter will seek to share humorous content. Therefore, health authorities may consider sharing content that humanises an infectious disease outbreak in order to counteract dehumanising humour during this time.

H.5 Sarcasm and Irony

There was overlap between the tweets coded as conveying sarcasm and those coded in other themes. Sarcasm and irony can be difficult to detect in day-to-day interactions, and on Twitter, where tweets may lack context, it can be even more difficult. Sarcasm is defined as:

> "the use of remarks that clearly mean the opposite of what they say, made in order to hurt someone's feelings or to criticise something in a humorous way" (Sarcasm. (n.d.). In Cambridge Dictionary)

Irony is defined as:

> "a situation in which something which was intended to have particular result has the opposite or a very different result" (Irony. (n.d.). In Cambridge Dictionary).

The irony definition above was used to indicate those tweets which may mean the opposite of what they express. Tweets in this sub-theme were sarcastic, ironic or contained elements of sarcasm and irony without satisfying the definition in full. In other words Twitter users did not necessarily have to insult or criticise other users to be placed in this category.

Twitter users made claims that were sarcastic and tweets in this category overlapped with tweets from other categories, and it was found that there was cross-over to other themes. Illustrative examples of such tweets include:

If you have sex with me, you will be immune to Swine Flu (H5.1)

I don't have Swine Flu by I do have bacon fever! (H5.2)

Omg, Swine Flu is spreading on Twitter! (H5.3)

'Drinking Budwiser – that will keep the Swine Flu away! (H5.4)

It must be noted that there may be differences in how researchers interpret these tweets, since sarcasm and irony can be difficult to detect and pinpoint within tweets. The tweets in this category could have a sarcastic tone, without fully matching the definition of sarcasm. Users often made claims that were ironic. These findings may indicate that a subset of Twitter users did not take the swine flu outbreak seriously or attempted to lighten the situation. Tweets which expressed sarcasm could potentially provide insight into some of the topics that were being discussed at the time. For example, in tweet *H5.3,* the user tweeting links swine flu to bacon, and at the time many Twitter users were unsure of whether bacon and pork produce could transmit swine flu.

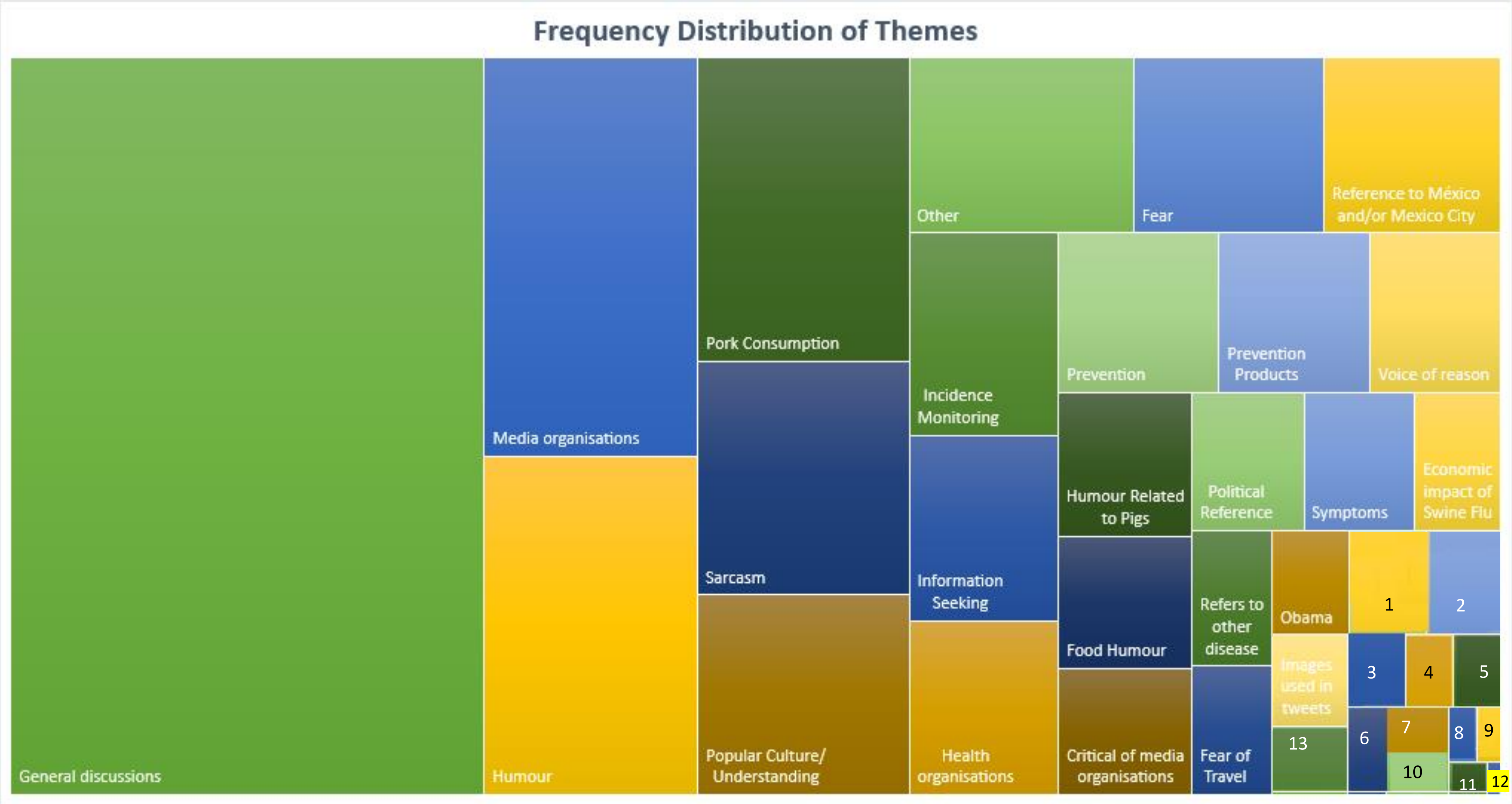
4.8 Results from Quantitative Analysis

A number of quantitative analyses were conducted on the data and are presented in this section. This section also reports on a number of word clouds that were generated. Word clouds are a useful way to see the different types of vocabulary conveyed by Twitter users, and can also aid the interpretation of themes as they visually display the range of word choices and vocabulary used by Twitter users.

4.9 Frequency Distribution of Themes

Figure 4-3 below illustrates the sub-themes that were described in this chapter in relation to their relative frequency of occurrence.

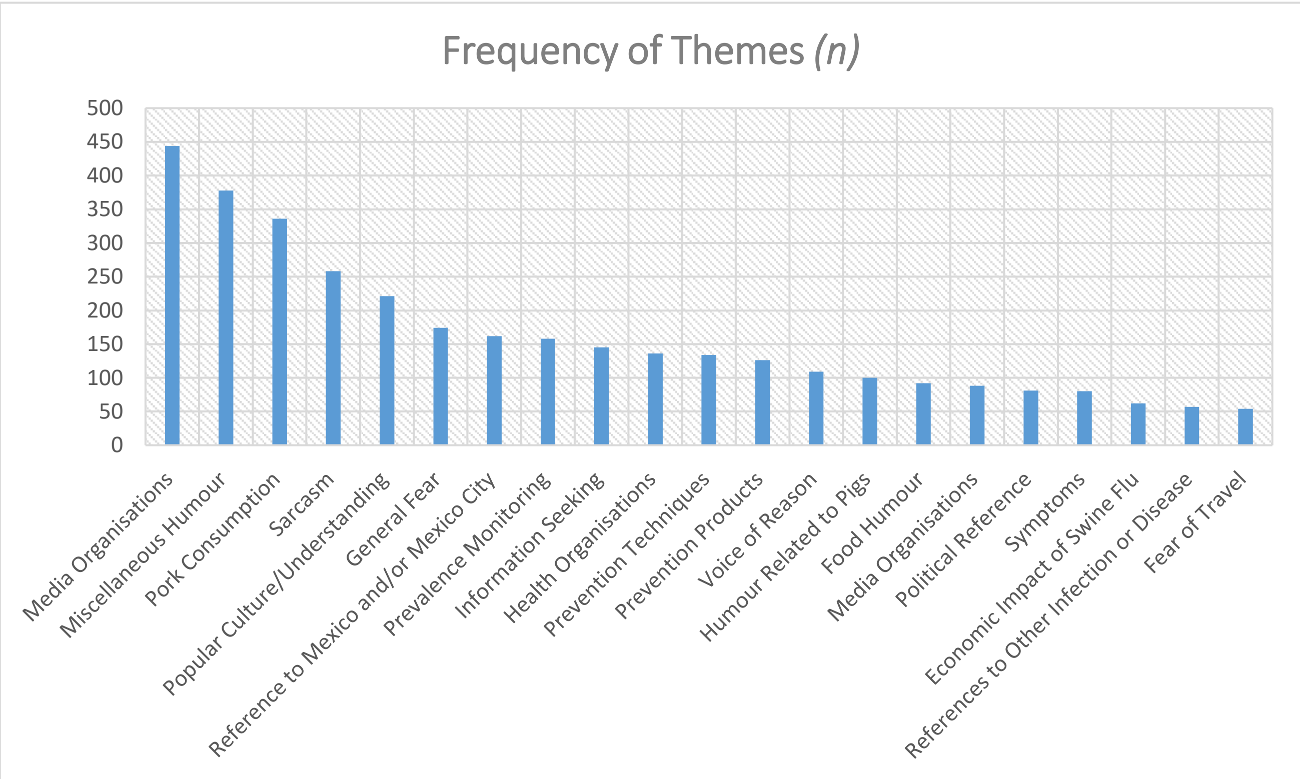
Figure 4-3 Frequency distribution of themes (swine flu)



*Notes
1. Reference to Mexicans
2. Resources
3. Transmission
4. Speculative Diagnosis
5. Nervous Humour
6. Anger
7. Medication
8. Worry
9. Critical of Media Organisations
10 Frightening Scenarios
11. Reference to Borders
12. Unfollowing users

A tree map (Figure 4-3) was selected in order to represent the data as it visually displays the different topics users were discussing and it displays these discussions relative to one another. It is interesting to note, from examining the tree map above, the number of different discussions related to swine flu taking place during this time, and which had differing levels of frequency. The theme General discussions was the broadest and is composed of a collection of different topics. Figure 4-4 below displays the frequency of sub-themes that had over 50 coding references excluding the theme of general discussions and other.

Figure 4-4 Frequency of themes



The theme of *General discussions* was removed in order to generate the bar chart presented above because it would have overshadowed the other themes. The sub-theme of *Other* was also removed because it was not useful in aiding the comparison among themes with content. The four most frequently occurring sub-themes consisted of: Media Organisations (444; 5.8%); Humour (378; 4.9%); Pork Consumption (336; 4.4%) and Sarcasm (258; 3.3%). It is interesting to note the two themes *Media organisations* and *Critical of media coverage* both mention the media. This could indicate that there were several discussions surrounding the media during the pandemic.

Media organisations were disseminating information related to swine flu during this time. However, the mainstream media became a subject of discussion as users began to hit back at news coverage. By running sensationalist headlines, news organisations may have received increased readership and viewership, which would have a positive effect on their revenue. Health authorities, therefore, may have to remain vigilant during outbreaks of infectious diseases as the potential for misinformation and sensationalist headlines could potentially arise.

It can also be seen that the themes of fear and fear of travel potentially suggest that users were feeling afraid due to the swine flu outbreak in general, and had specific travel-related fears. Health authorities may wish to disseminate information in relation to stress reduction strategies during infectious disease outbreaks. This section has examined the themes that emerged from the nodes, and the next section examines the distributions of nodes that formed from the analysis of tweets.

4.10 Popular Content

In 2009, when the data for this case study were retrieved, Twitter did not store data such as number of retweets, and location data associated with tweets (GNIP, 2017). Therefore, it has not been possible to report this information in this chapter. In order to identify popular tweets, DiscoverText's clustering algorithm was used to group all identical tweets (214,784) from the two-day period April 28th to April 29th which were over five characters in length. The table below provides an overview of the most frequently occurring tweets within the dataset.

Table 4-14 Most frequently occurring tweets

Rank	Tweet illustration	Tweet Frequency (%)
1	There is swine flu hysteria over just 10 deaths	395 (0.18%)
2	You can track the outbreak of swine flu by using Google Maps	360 (0.16%)
3	There is complete coverage of swine flu news at swine-flu.alltop.com	270 (0.12%)
4	CDC Emergency: it is not possible to get swine flu from eating pork	164 (0.07%)
5	How concerned are you about the number of swine flu cases in the US?	116 (0.05%)
6	I am unable to tell if I have the swine flu or whether I need a fresh air cartridge	104 (0.04%)
7	This swine flu cartoon is mocking us all	102 (0.04%)
8	CDC warns against all non-essential travel to Mexico	99 (0.04%)
9	Swine flu is spreading to the Middle East	93 (0.04%)
10	Twitter is catching swine flu	88 (0.04%)

It can be seen from the table above that the most frequently occurring tweet is downplaying swine flu by noting that there has been hysteria over just ten deaths. This tweet expresses a view which was found in theme *C.4 Voice of Reason.* Additionally, people might have been using Twitter to address their information needs in order to assess their perceived susceptibility of the swine flu outbreak. This could explain why a tweet related to tracking swine flu via Google Maps that appears among the most frequently occurring is because it allows users to assess their perceived severity of the disease. This tweet expresses the view found in theme *B.2 Prevalence Monitoring.* Similarly, tweets that were shared most frequently could be said to address an information need, and so were particularly popular among the public. For example, tweet 4 and 8 from Table 4-14 are from the CDC and address potential information needs. Tweet 5 seeks information based on whether people are concerned with the number of swine flu cases, and would fall in theme *C.2 Information Seeking.* There are also some humorous tweets such as tweet 5, 7 and 10, which express views found in the *H. Humour and Sarcasm theme.* The table below displays some of the most frequently used hashtags from all tweets (214,784) from the two day period, April 28th to April 29th 2009.

Table 4-15 Most frequently occurring hashtags

Hashtag	**Number (%)**
#swineflu	32,043 (15.4)
#swine	1,913 (0.9)
#influenza	976 (0.5)
#tcot	774 (0.4)
#flu	763 (0.4)
#h1n1	745 (0.4)
#news	654 (0.3)
#mexico	366 (0.2)
#pandemic	340 (0.2)
#fb	334 (0.2)

As presented in the table above, the most frequently occurring hashtags were that of #swineflu (32,043), #swine (1,913), and #influenza (976). These hashtags could be popular via genuine usage, or they could appear here due to link-baiting, i.e. that organisations or individuals would use the hashtag to try to increase visibility of their own tweets which may contain links. It is interesting to observe the *''tcot'* hashtag which stands for 'Top Conservatives on Twitter' and was set up by conservative activists as a communication back-channel to share and discuss ideas (Beutler, 2009) appears among the most popular used hashtags.

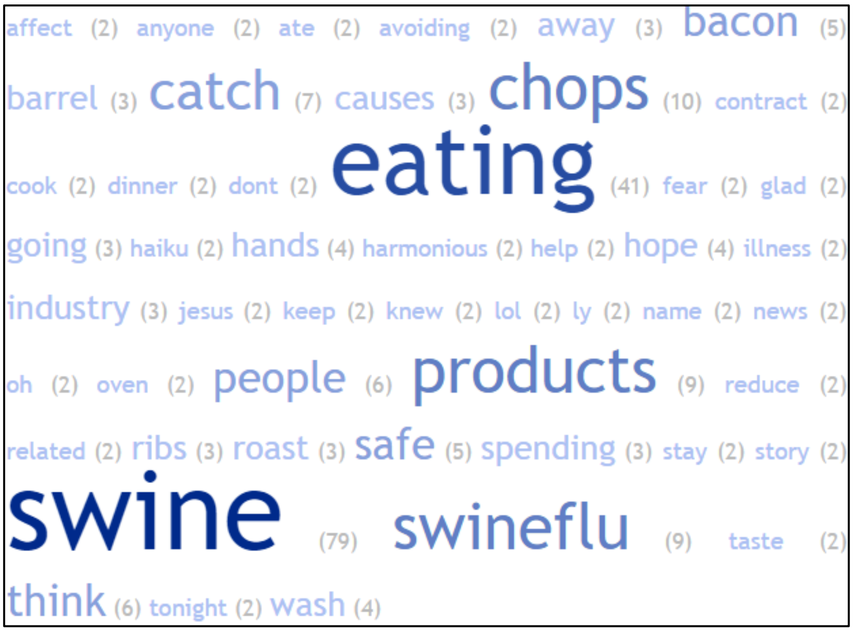
4.11 TAG Clouds of Sub-Themes

This section examines tag clouds from certain sub-themes that were generated in order to gain insight into the vocabulary of Twitter users. The size of the words are related to the frequency of occurrence, i.e. if a word occurs more frequently, it will appear in larger font within the word cloud. The word cloud is limited to 50 of the most popular words, and similar words will be grouped together (e.g. the words 'flu' and 'Flu'). The application used to generate the TAG clouds was Tag Crowd (Steinbock, n.d.). The TAG cloud also contains the frequency of occurrence associated with each of the words. Stop words, i.e. those words that were not included because they occur so frequently in everyday usage, were removed from the tag clouds, they included words such as: 'com' 'http' 'rt' 'tinyurl' and 'www'. Moreover, words such as 'swine flu', 'swineflu', 'swine' and 'flu' were also removed. This is because these words appear across sub-themes. Word clouds are useful as they provide new insight into the vocabulary utilised by the general public. The tag clouds below were created by extracting tweets coded in their respective nodes.

Figure 4-5 Tag cloud of tweets which were mentioning the media being critical

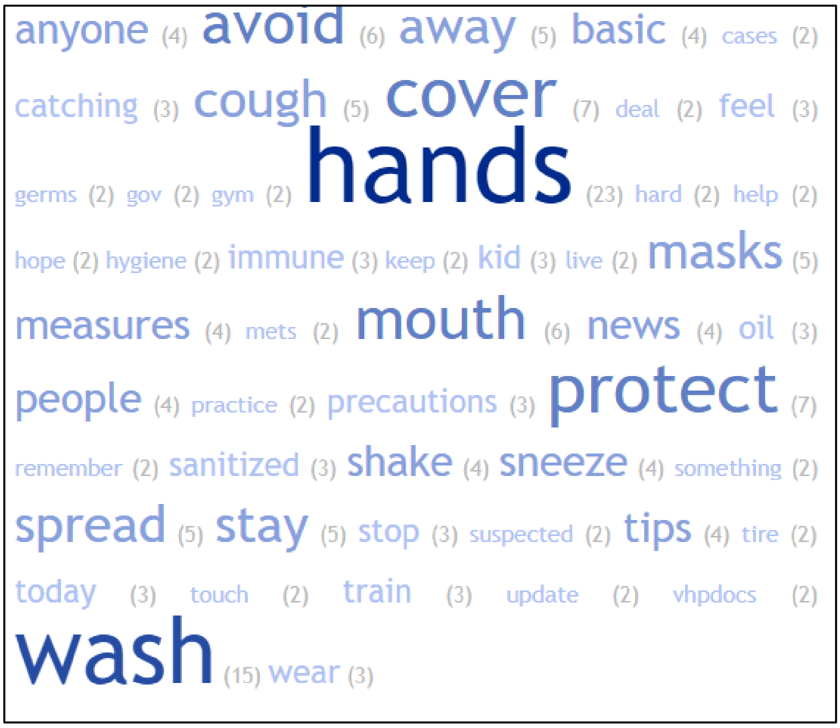
It is interesting to see words such as such as ‘blowing’ and ’proportion’ appear as they were used to form the phrase *‘media is blowing swine Flu out of proportion’*. Other words are that of ‘panic’ as users noted the media were causing users to panic, and there were also words such as ‘sick’ and ‘hearing’, as users would note that they were *‘sick of hearing of the news’* related to swine flu. Stop-words consisted of: coverage, flu, http, media, news, rt, swine, and swineflu.

Figure 4-6 TAG cloud of tweets which were mentioning pork



The most frequently occurring words are that of ‘eating’, ‘chops’ and ‘products’. These words may have been used to discuss and question whether consuming pork would cause the transmission of swine flu. Other words such as ‘bacon’ and ‘pork’ appear as users tweeting at the time would question whether eating such products could transmit swine flu. Stop words included: bc, com, flu, http, pork, rt, and tinyurl.

Figure 4-7 Word cloud of tweets which mention prevention

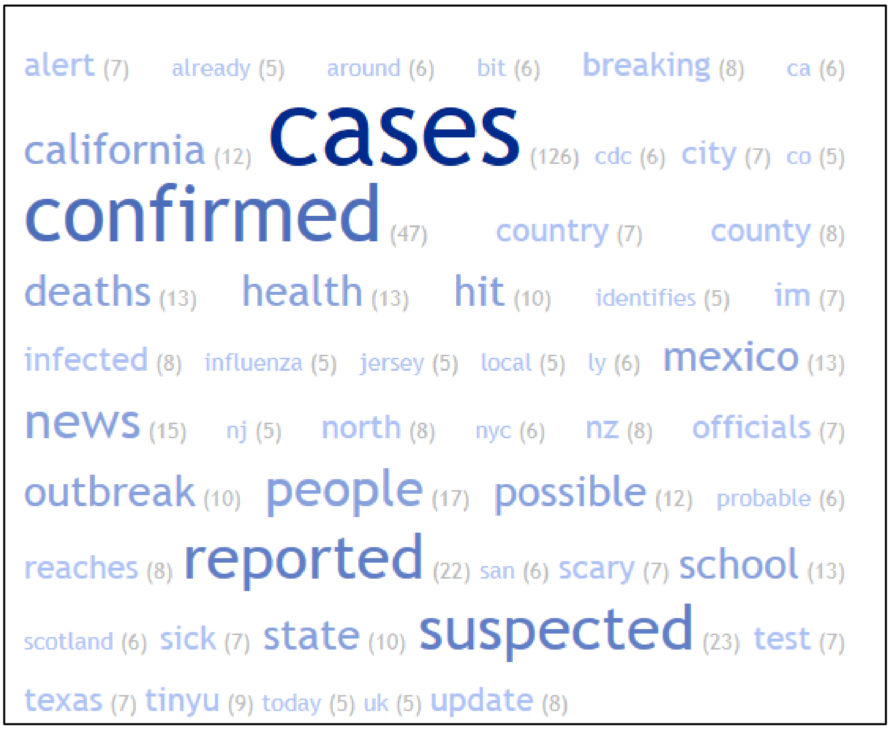


The word cloud above displays a range of vocabulary that was used by Twitter users when discussing the topic of how swine flu could be prevented. We can see that the word '*hands*' appears in order to form the sentence '*wash hands'.* Other words that may be of interest include '*mouth'* as Twitter users may have been referring to the act of covering the mouth, i.e. to protect ofrom swine flu during coughing or sneezing. Stop words included: best, com, cuz, don't, flu, htm, http, prevention, rt, swi, swine, swineflu, tinyurl, ur, and, www.

Figure 4-8 Word cloud of tweets which mention symptoms

Words such as 'headache' and 'throat' appear in the word cloud. This could potentially suggest that users associated the symptoms of headaches and sore throats with swine flu. The words generated as occurring frequently may be of potential interest to health authorities that disseminate information during this time period as users who were suffering from headaches and sore throats would ask whether they had swine flu. Stop words included: 'andersoncooper', 'Brandon', 'com', 'flu','http', 'im', 'oh', 'rt', 'swine', 'swineflu', 'symptoms', 'tinyurl', and 'ur'. The stop words 'andersoncooper' refers to Anderson Cooper and 'Brandon' refers to Brandon Keim who are both journalists.

Figure 4-9 Word cloud of tweets which were monitoring the prevalence of swine flu



The words 'cases' and 'confirmed' appear among words within the word cloud as this refers to the term 'confirmed cases'. There are also mentions for locations such as 'Mexico' and 'California' as these were locations affected by swine flu during the outbreak.

4.12 Validity and Reliability

4.12.1 Test Retest Reliability

Test-retest reliability (as outlined in section 3.15) is measured by having the same individual code a dataset after a period of time in order to compare results. It is said to be able to measure the reproducibility of a set of results (Litwin, 1995). The test-retest reliability was measured by using Cohen's kappa coefficient. The author who coded the initial study (WA), also re-coded a sub-set of data after a three-month period in order to assess test-retest reliability. The percentage agreement was 99.94%, and κ =0.85, which is at a substantial level (McHugh, 2012).

4.13 Intercoder Reliability

A Doctoral student with Masters level education from the Information School was provided with verbal instructions, and trained to code 300 tweets in order to produce an intercoder reliability statistic. Intercoder reliability percentage agreement was 99.96%, and $\kappa = 0.53$ which is at a moderate level (McHugh, 2012). Intercoder reliability was outlined in the section 3.15.

4.14 Identifying Influential Twitter users

It is possible to identify Twitter users who were being mentioned in association with swine flu related keywords with high frequency. The method used to calculate InDegree and OutDegree is known as social network analysis (as explained in methodology Chapter 3 section 3.7.1). This study made use of Gephi (Gephi, 2014) in order to identify highly-ranked Twitter users by InDegree and OutDegree from a sample of the network in order to provide insight into the types of tweets these users were sending. The study examined the Twitter network of swine flu and sampled the network from 30th of April. (09:00-12:00). The network contained 21725 vertices, i.e. number of users within the network, and there were a total of 7165 edges which refers to the connections between users.

4.14.1 Examining influential Twitter users by InDegree

Table 4-16 below outlines a rank of Twitter users who received a number of inbound connections alongside the keywords: ‘swine flu’, ‘#SwineFlu’, and ‘H1H1’. The metric of InDegree looks for users who were mentioned frequently alongside these words, used to capture the data. The column labelled *Node* is a label assigned to each of the Twitter users, the InDegree as described above is the number of mentions alongside the keywords, and the account description proves detail on the account type. These headings also apply to Table 4-17.

Table 4-15 Ranking Twitter users by InDegree Sample from last three hours

Node	InDegree	Account Description
A	168	Citizen, Entrepreneur and Technology Investor
B	111	CNN Breaking News
C	91	Celebrity, Olivia Munn
D	86	NBC's Breaking News
E	14	Parody Swine Flu Account

Node A - Citizen, Entrepreneur and Technology Investor

The highest ranked Twitter user by InDegree was an Entrepreneur and Technology Investor from New York who sent a number of sarcastic and humorous tweets related to the swine flu outbreak. These tweets were shared and retweeted by a number of Twitter users. For example, a tweet that was shared widely occurred when the user noted that they would rename 'swine flu' to 'bacon flu' as that would sound less scary and 'slightly tasty'. The user also tweeted a link to an article that they had published which was critical of the false information on Twitter.

Node B – CNN Breaking News

The second most highly ranked Twitter user was the Twitter account entitled *CNN Breaking News*, an account which provides news to Twitter users. The tweets sent by CNN across this time period related to the news that the WHO had raised the swine flu alert level to level 5, and that a 23 month-old Texas child had been confirmed as the first US swine flu death.

Node C – Olivia Munn

The third most highly ranked Twitter user was a celebrity, Olivia Munn, who tweeted on April 29th 2009 that she was getting sick, was nauseous, had a fever, but was not pregnant. Olivia did not tweet that she had swine flu, and none of her tweets mentioned swine flu during this time period. However, at least 71 unique users replied to Olivia suggesting that the reason she may be feeling ill may be due to swine flu, others taunted and insulted her, and one particular Twitter user tweeted: *'hope you get Swine Flu'*, although other Twitter users hoped that she did not have swine flu. As a consequence of these interactions, Olivia's Twitter account received a large number of inbound connections from Twitter users. During this time there was highlighted interest surrounding swine flu and, therefore, Twitter users may have thought that Olivia was referring to swine flu.

Node D – NBC Breaking News

NBC's breaking news account appears to have a high InDegree because the account was providing news updates in real time related to the swine flu outbreak, and consequently it had sent a number of tweets which used the keyword swine flu. In particular, during this time period, the account shared news that the US state department had issued a travel alert to Mexico because of the swine flu outbreak, and Twitter users were sharing this post. The Twitter account of NBC also received a large number of inbound mentions because Twitter users criticised the account for providing too much swine flu content.

Node E – The Swine Flu

A parody Twitter account titled *The Swine Flu* received many mentions by Twitter users who referenced the existence of the account and/or shared a joke and tag the account. During this time period, the account tweeted that the movie '28 Days Later' was a source of inspiration for the Swine Flu, and that swine flu was thinking of mutating. As a consequence of this, the replies to the account were mostly humorous and these were retweeted.

4.14.2 Examining top-ranked Twitter users by OutDegree

Table 4-167 Ranking Twitter users by OutDegree

Node	OutDegree	Account Description
A	21	Member of Public
B	17	News Aggregation account
C	15	Account which tags top tweeters
D	14	Telework Twitter account
E	14	Member of Public

Node A – Member of the Public

This Twitter account was from a member of the public who was replying to Twitter users frequently. However, the majority of tweets that were sent were of a humorous and sarcastic nature. The account had a very high OutDegree as the Twitter user had mentioned swine flu in their tweets with high frequency.

Node B – News Aggregation Account

The Twiter account labelled as node *B* was of a Twitter account which provides aggregated news to its followers. The account was retweeting with high frequency and some of the nodes that were retweeted by this account belonged to organisations such as CBS News, BBC World, and the Miami Herald.

Node C – Account which Tags Top Tweeters

This Twitter account tagged users that according to the account were ranked as most influential. The motto of the account was that they would help people find other people on Twitter who were influential. It notes that the account helps locate people on Twitter.

Node D - Telework Twitter Account

A Twitter account which was related to the concept of work-shifting tweeted with high frequency during the swine flu outbreak). The account attempted to promote the idea of teleworking, working from home and work-shifting. In the context of the swine flu outbreak, it appeared that the account was informing users that teleworking would reduce risk of developing swine flu.

Node E - Member of public with interest in programming

A Twitter account from a member of the public, which identified as having an interest programming languages such as python, sent a number of humorous tweets related to the swine flu outbreak.

4.15 Evidence of themes across the outbreak

One of the questions that can be asked of the study is whether the themes identified in the 2-day period are applicable, i.e. scale to other activity on Twitter. It was not feasible to obtain the complete data on the swine flu outbreak because of the high cost of obtaining historical Twitter data and managing it. However, one method that can be used to search all of Twitter is Twitter's advance search feature. In this study, Twitter's advance search feature was utilised in order to find tweets from across the pandemic from when Google Trends showed there to be an increased interest in web search queries, and this was from January 2009 until November 2009. Themes were selected if they were not specific to the two-day time period, were not reported in previous literature, and ere searchable using the keywords. By running a number of searchers on Twitter's advance search it was found there were tweets from across the pandemic (May to November) that expressed opinions which were identified in the following themes: A.1 General Fear, C.4 Voice of Reason, C.6 Name Discussion, D.4 Media Organisations Critical, F.2 Reference to Mexicans, and G.1 Pork Consumption. Appendix 3 contains examples of tweets from the beginning middle and end of each month from May 2009 to November 2009 associated with the themes above.

Table 4-17 Themes across the outbreak

Themes	May	June	July	Aug	Sep	Nov
A.1 General Fear	x	x	x	x	x	X
C.6 Name Discussion	x	x	x	x	x	X
D.4 Media Organisations Critical	x	x	x	x	x	x
F.2 Reference to Mexicans	x	x	x	x	x	x
G.1 Pork Consumption	x	x	x	x	x	x
H.3 Popular Culture/ Understanding	x	x	x	x	x	x

These results provide evidence that certain themes were occurring throughout the outbreak and were not specific to the 2-day period. This means that users would express fear in their tweets, discuss the name given to swine flu, critique the media, refer to Mexicans, discuss whether pork was safe to consume, and refer to popular culture throughout the outbreak period.

4.16 Discussion of Results

4.16.1 Key Findings

This study found that the information which was shared on Twitter during this time period revolved around eight key themes:

- Emotion and feeling (4.4%)
- Health-related information (10.6%)
- General commentary and resources (43.0%)
- Media and health organisations (11.8%)
- Politics (2.2%)
- Country of origin (3.7%)
- Food (7.5%)
- Humour and/or sarcasm (17%)

It is important to note that previous studies examining swine flu discussions on Twitter (Kostkova, Szomszor and St. Louis 2014; Signorini, Segre and Polgreen 2011; Chew and Eysenbach, 2010) did so with the aim of validating the potential of Twitter as an early warning system and as a tool to extract public sentiments, and none of these studies used an in-depth qualitative approach to analyse data. The aim of the present research was to study the content of tweets using thematic analysis and to examine the types of conversations that were taking place. In doing so, the study developed new knowledge on the types of information that were shared by users and which had not been reported in previous literature. In Chapter 7 Discussion, Section 7.3.9, the results of this study are compared to previous research on this topic, and it was found that at least 23 new themes emerged in this present study that, to the best of the author's knowledge, have not been reported by previous empirical studies. Hence, this study developed new knowledge related to the potential of popular culture influencing how Twitter users were understanding and speaking about swine flu. For example, some Twitter users made references to zombies and to a zombie apocalypse, and these tweets were categorised within the popular culture sub-theme (3.9%). Another interesting finding was that Twitter users were using the platform to seek information (2.5%) during the outbreak by tweeting open-ended questions, as reported in a previous study (Chew and Eysenbach, 2010).

This study also developed new knowledge on how Twitter users would refer to Mexico and to people of Mexican descent. For example, a number of users blamed Mexico for the outbreak of the swine flu virus, and constructed negative tweets towards people of Mexican descent, although not all tweets were negative in sentiment. These references accounted for 3.7% of the discussions. This study also found that Twitter users took an interest in tracking and monitoring the spread of swine flu (2.8%), and the term "citizen disease surveillance" to describe this behaviour was defined. This study developed new knowledge on how Twitter users shared information on prevention products (2.2%), for example, facemasks. Some Twitter users asked other users to remain calm, and a sub-theme downplaying the risk of swine flu (1.9%) emerged (*C.4 Voice of Reason*) from the data.

When Twitter users were identified as influential by InDegree, users who were influential were either accounts sharing humorous content and/or those which providing breaking news. When the metric of OutDegree, i.e. users tweeting the most, was applied, it appeared that users who were tweeting regularly were sharing spam. It was also found that the informal name of 'swine flu' given to the H1N1 virus lead Twitter users to incorrectly believe that pigs and pork meat could host and transmit the virus. This belief appeared across other themes and sub-themes such as G.*Food, C.8 Images Used or Referenced in Tweets*, and *B.3 Prevention.* The view most likely influenced the beliefs people held, specifically that avoiding pork would put them at decreased risk of developing swine flu. This was a serious issue at the time, and a campaign also started later in the year (August 2009), backed by the pork industry and farmers, aiming to promote the use the medical term of H1N1 rather than swine flu among the mainstream media and general public (Pork Magazine, 2009). Previous studies have noted the potential of the mainstream media to influence conversations on Twitter (Chew and Eysenbach, 2009) because users will share these articles on the platform. In the present study, it was found that the mainstream media were mentioned frequently (7.7%), often critically (1.5%). The results of this study will be further discussed and synthesised in Chapter 7: Discussion.

4.17 Health Belief Model

The theoretical framework of the Health Belief Model was applied throughout the analysis of tweets. The Health Belief Model purports that, if the perceived threat of a disease is high, then people are more likely to change their behaviour to avoid developing the disease. The perceived threat of a disease will be influenced by a number of factors such as whether a person is likely to get the disease as well as the seriousness of the disease. Additionally, a number of demographic variables as well as the personality and intelligence of a person will also affect the decision-making process when evaluating whether a disease has a high perceived threat, susceptibility and seriousness, and when weighing up the perception of barriers and benefits in taking a certain action. For instance, in the case of swine flu, the likelihood that someone would change their behaviour might be very high as the perceived seriousness of swine flu is high. Increased media coverage may lead to an increase in the perceived seriousness of a disease and, as a result, the general public may be more likely to alter their behaviour. However, a risk of increased media coverage is that users may begin to stigmatise those whom develop swine flu, and people from specific regions and of particular races may be targeted. It was found that the Health Belief Model was a useful theoretical framework to aid the understanding and interpretation how Twitter users were responding to the swine flu outbreak. One of the limitations of the model is that it may appear to over-simplify human behaviour. The utility of the Health Belief Model is summarised in Chapter 7 Section 7.5 which provides an outline of where the model was specifically utilised and its usefulness. The Health Belief Model will be further applied in Chapter 8 which compares the 2009 swine flu outbreak to the 2014 Ebola outbreak.

4.18 Limitations

This study analysed tweets from a two-day period, hence the findings are not applicable to all Twitter activity related to swine flu during the 2009 pandemic. However, section 4.15 did provide some evidence that certain narratives identified in this study were expressed by users across the outbreak. When analysing tweets and Twitter data, a known limitation among the research community is that Twitter data does not reflect the offline population or even the online population to an extent. However, the aim of this study was to examine content that was shared on Twitter from a public health informatics perspective, rather than to generalise findings to the offline population. As was noted in the results presented in this chapter, certain age groups (i.e. younger people) and geographical locations (i.e. the US) may be over-represented (Pew Research Center, 2016). It must also be noted that people might behave differently online compared in real life situations, and people may behave differently on Twitter compared to what they might do in a research interview: for example, views in tweets may be exaggerated as a reaction to an event. However, one strength of analysing Twitter data is that it avoids the potential of interviewer bias. Due to the qualitative nature of the study, there may be slight variations in the themes and sub-themes that could have emerged if the study was repeated by another researcher. This is because tweets may be interpreted differently by different researchers and this can affect the types of themes and sub-themes identified. Measures were taken to ensure the validity and reliability of results, such as by performing intercoder and test-retest reliability. Moreover, by examining English-language tweets and excluding tweets in other languages such as Spanish, the study is limited because it may not have captured the voices of those who whom were affected in Mexico where the virus initially emerged as they may have been tweeting in Spanish.

4.19 Summary

In summary, a number of interesting and original results across the themes and sub-themes were identified from the data, for instance, the amount of health-related information that was shared on Twitter, and to how citizens actively engage with the platform for prevalence monitoring. It also emerged that swine flu caused Twitter users to express several emotions during the outbreak, including nervous humour. Moreover, it was interesting to note that the term ‘swine flu’ caused confusion among some Twitter users. Potentially, it may have been better for news organisations to have referred to the swine flu virus in other terms to prevent citizens from forming a causal link to pigs, to avoid offending certain ethnic groups, and to reduce the occurrence of misinformation and the economic impact that swine flu may have had on farming. Moreover, it was found that Twitter users would refer to popular culture in order to understand the outbreak.

Further discussion will take place in Chapter 6, which discuss the results of the individual case studies in comparison to one another. These findings will be of interest to health dissemination and public health literacy organisations in order to be in a better-informed position when disseminating information on Twitter. The methodology of filtering a large quantitative dataset used in this study may serve as a useful guide for health organisations that wish to sample and analyse tweet content during the height of an outbreak. Moreover, the results presented in this chapter are likely to inform public health strategies for future infectious disease outbreaks, as well as current outbreaks such as the Middle East respiratory syndrome (MERS), the Zika virus, and H5N1.

Chapter 5 Ebola Study

5.1 Introduction

This chapter describe the results of a study that was undertaken in which data were generated from Twitter and a thematic analysis was undertaken. Eight prominent themes that emerged from analysing data related to the Ebola outbreak from the 29th to the 30th September 2014. The chapter first provides some background information on the Ebola outbreak and places it into context (section 5.2 and section 5.3), then it reports on a number of findings which emerged from the analysis of data (section 5.4). Where possible, the Health Belief Model was applied to understand why Twitter users acted and communicated in certain ways.

5.2 Background

The Ebola Virus Disease, previously named the Ebola haemorrhagic fever, is a very severe illness with an average fatality rate of 50% (WHO, 2015). Ebola was first identified in 1976 when two outbreaks occurred simultaneously in Nzara (Sudan) and in Yambuku (Democratic Republic of Congo) (Briand *et al.*, 2014; WHO, 2015). The most recent outbreak of Ebola began in Guinea in December 2013 (Baize *et al.*, 2014; Briand *et al.*, 2014) and then spread to Liberia and Sierra Leone in West Africa. The first case of Ebola outside of Africa was reported in the United States on 19th September 2014, and the first diagnosed infection occurring outside of West Africa was reported on 30th September 2014 (WHO, 2015).

The 2014 outbreak of Ebola is the largest ever recorded, and has affected more people and taken more lives than all of the other Ebola outbreaks combined (Briand *et al.*, 2014; WHO, 2015). On the 8th of August 2014 (33 weeks into the outbreak), the WHO declared the epidemic to be a Public Health Emergency of International Concern (PHEIC). This decision was not taken lightly (Briand *et al.*, 2014), because the PHEIC is an instrument of the 'International Health Regulations (IHR)' an agreement which is legally binding made by 196 countries on containment of major international health threats (Briand *et al.*, 2014).

The World Health Organisation (2015) stated that Ebola is transmitted to humans via wild animals and can be spread via human-to-human contact. Ebola is not airborne and it is unlikely that a mutation will led to its becoming airborne (Centre of Disease Control, 2015). Ebola is transmitted to humans via blood and secretions of an infected person or via surfaces which contain these fluids (WHO, 2015). Ceremonies of those being burred where those who are mourning have contact, which is direct with the body of a deceased person, may also aid the transmission of Ebola (WHO, 2015). The WHO (2015) reported that Ebola has an incubation period of two to 21 days and that humans will only become infectious after symptoms have developed. The initial symptoms include sudden fever and fatigue, muscle pain, headache, and sore throat, which are then followed impaired liver and kidney functions and both internal and external bleeding (WHO, 2015). Symptoms can also include diarrhoea, skin rash, and vomiting (WHO, 2015). Diagnosing Ebola is challenging, as it is difficult to differentiate it from other infectious diseases. However, by taking blood or stool samples from a patient and conducting diagnostic tests, it is possible to identify it. The samples provided by patients are a biohazard, hence laboratory testing must take place in biological containment conditions (WHO, 2015). There is no treatment for Ebola; therefore, supportive care-rehydration via oral and intravenous fluids is provided to patients (Briand *et al.*, 2014; WHO, 2015). For prevention and control, the WHO (2015) suggested that successful community engagement is one of the ways of controlling outbreaks. Raising awareness for Ebola, including protective measures that people can take, is effective in reducing transmission (WHO, 2015). The WHO (2015) recommended that health workers who are in contact (within a metre) of Ebola patients should undertake measures to prevent contagion, such as face protection (face shield, medical mask, goggles) and a clean non-sterilised gown and gloves.

5.3 Summary of Key Events

It is important to understand the events that unfolded during the Ebola outbreak, as Twitter users may have been influenced by news events and/or shared news from articles they had been exposed to. Tables 5-1 and 5-2 provide a brief overview of events that took place from the 24th August 2014 to the 19th September 2014. This study examined tweets from the 29th to the 30th September 2014, however, it is important to examine events that took place slightly before and after this time period as context. The tables below were produced using information provided in an article by the Guardian (Jansen, 2014).

Table 5-1 Summary of events taking place from 26th to 29th August 2014

Date	24th August 2014	26th August 2014	27th August 2014	29th August 2014
Event	• A British nurse William Pooley is brought back to UK after contracting Ebola in Sierra Leone.	• Reports that Ebola may have been diagnosed in Democratic Republic of Congo.	• Medecins Sans Frontieres state international response to Ebola is slow and negligent.	• Ebola is reported in Senegal and there are riots in Guinea as rumors circulate that health workers are transmitting the virus to locals.

Table 5- 2 Summary of events taking place from 5th September to 29th September 2014

Date	5th September 2014	16th September 2014	26th September 2014	29th September 2014
Event	• 2,100 deaths and 4,000 infected according to WHO,	• Obama announced US to send 3,000 troops to West Africa, and notes that the outbreak is out of control.	• 3,091 deaths, and 6,574 suspected cases according to WHO. Liberia, Guinea, and Sierra Leone are the most badly affected.	• First person diagnose d in the US (Dallas).

Figure 5-1 Time series graph of Twitter activity for keyword Ebola for the period 28th and 29th September 2014

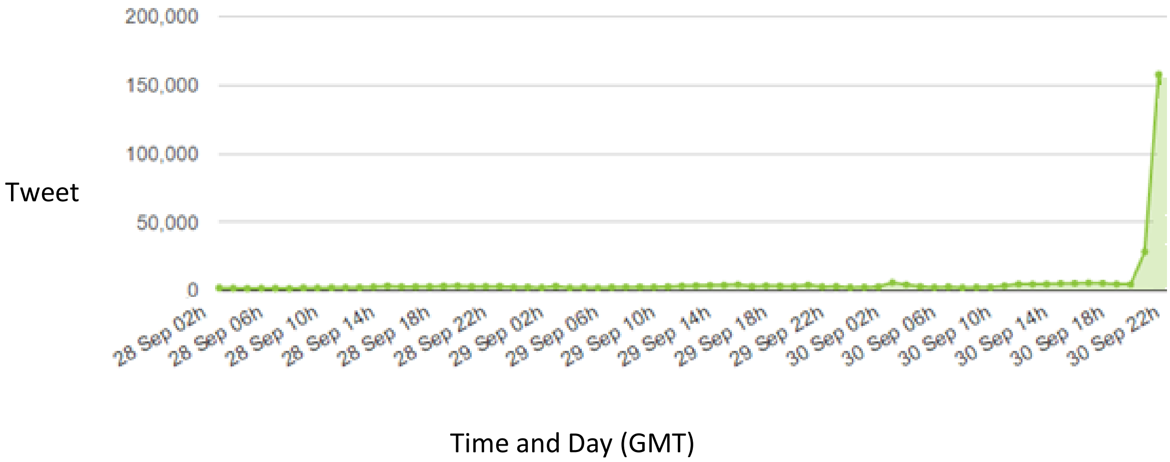


Figure 5-1 is a time series graph which displays the Twitter activity of the keyword 'Ebola' from the 28th to the 29th September 2014. After an initial phase of relatively little activity from 28th September until 29-th September, there is a peak of tweets on the 29th September 2014 which corresponds with the news that a person had been diagnosed with Ebola in the US in Dallas. At the highest peak, a total of 157,762 tweets were sent and received. Data was not selected from when there was the highest peak because the content appeared to revolve solely around a single case of Ebola in the US, as found in the pilot study (section 3.10).

5.4 Summary of Data Collection and Justification

The data collection procedures (as outlined in Chapter 3 Methodology section 3.14) are summarised below:

- The entire dataset that was retrieved in relation to Ebola consisted of 181,110 tweets from the time period of 28th to 29th September 2014.
- This interval of time was selected because Google Trends Data showed an increase in Web Search queries (Google, 2017).
- The keywords used to retrieve data were 'Ebola' (section 3.5.1 provides an overview of the challenges of retrieving data via keywords).
- This dataset was retrieved using Visibrain (n.d.), which is a licensed reseller of Twitter data.
- Using DiscoverText (n.d.), duplicate clusters (i.e. similar tweets) were removed, resulting in 102,852 tweets.

- Near-duplicate tweets were removed at a 60% threshold leading to a total of 56,948 single tweets.
- Finally, a simple random sample of 10% was taken (5,685 tweets), and entered into NVivo for coding using thematic analysis.
- A 10% simple random sample ensures that there is an equal chance of selecting each tweet and provides an unbiased representation of the entire dataset (Starnes, Yates, and Moore, 2010)

The next section provides an overview of the themes and sub-themes that emerged from the in-depth thematic analysis that was conducted on the dataset.

5.5 Summary of Platform Features

It is important to understand how Twitter was differed from its current form, as a number of feature changes were implemented in recent times. This will allow for future studies, as well as studies conducted in the past to be able to be compared to the present research (Ellison and Boyd, 2013). The changes that have been implemented to Twitter from 2009 to 2014 are described below:

- After 2014, a number of changes have been implemented at different times regarding how tweets are displayed on the Twitter timeline. For example, a new 'while you were away' feature was implemented after 2014 (Constine, 2014). This is displayed to users when they log into the platform and it shows them a list of recent tweets that they might have missed.
- After 2014, over different time intervals, Twitter provided more text for people to send tweets by not counting usernames, photos, videos, and polls within the 140-character limit (Wagner, 2017). Therefore, studies conducted after 2014 would be analysing tweets that could be longer in length.
- Twitter replaced the 'favorites' button which was in the shape of a star to that of a 'like' button which is in the shape of a love heart. This means that after 2014 Twitter users would have liked tweets rather than favourited them (Parkinson, 2015).
- After 2014, Twitter introduced a new feature called 'moments' which uses humans to curate popular tweets related to a breaking news event or an emerging story (Warren, 2015). This feature also allows users who are not logged into the platform to follow breaking news stories.

- In 2017 Twitter increased its text character limit from 140 to 280 characters (Telegraph, 2009) which means that future studies on infectious disease outbreaks may analyse tweets which may be longer in length than studies conducted prior to this.

5.6 Results from Qualitative Analysis

Eight prominent themes have emerged from the qualitative thematic analysis of tweets, and these are described below:

- Theme E: Emotion and Feeling. Tweets which expressed emotions towards the Ebola outbreak. These included anger, fear, fear of travel, praying, prayers and calls to God.
- Theme F: Health Related Information. Tweets which discussed medical concepts such as transmission, prevention, symptoms, and vaccines.
- Theme G: Significant News Stories**.** Tweets which referred to a specific news story which was significant during this time period.
- Theme H: General Commentary**.** Tweets which were discussing the Ebola outbreak in general terms.
- Theme I: References to Official Organisations. Tweets which referred to official organisations such as health organisations, and health charities.
- Theme J: References to a West African city or region. Tweets which would refer to regions or specific geographical areas in West Africa such as Sierra Leone.
- Theme K: Politics. Tweets which made reference to politics or a political figure or offered a critical view of governments.
- Theme L: Humour and/or Sarcasm. Tweets which refer to miscellaneous humour and sarcasm related to Ebola.

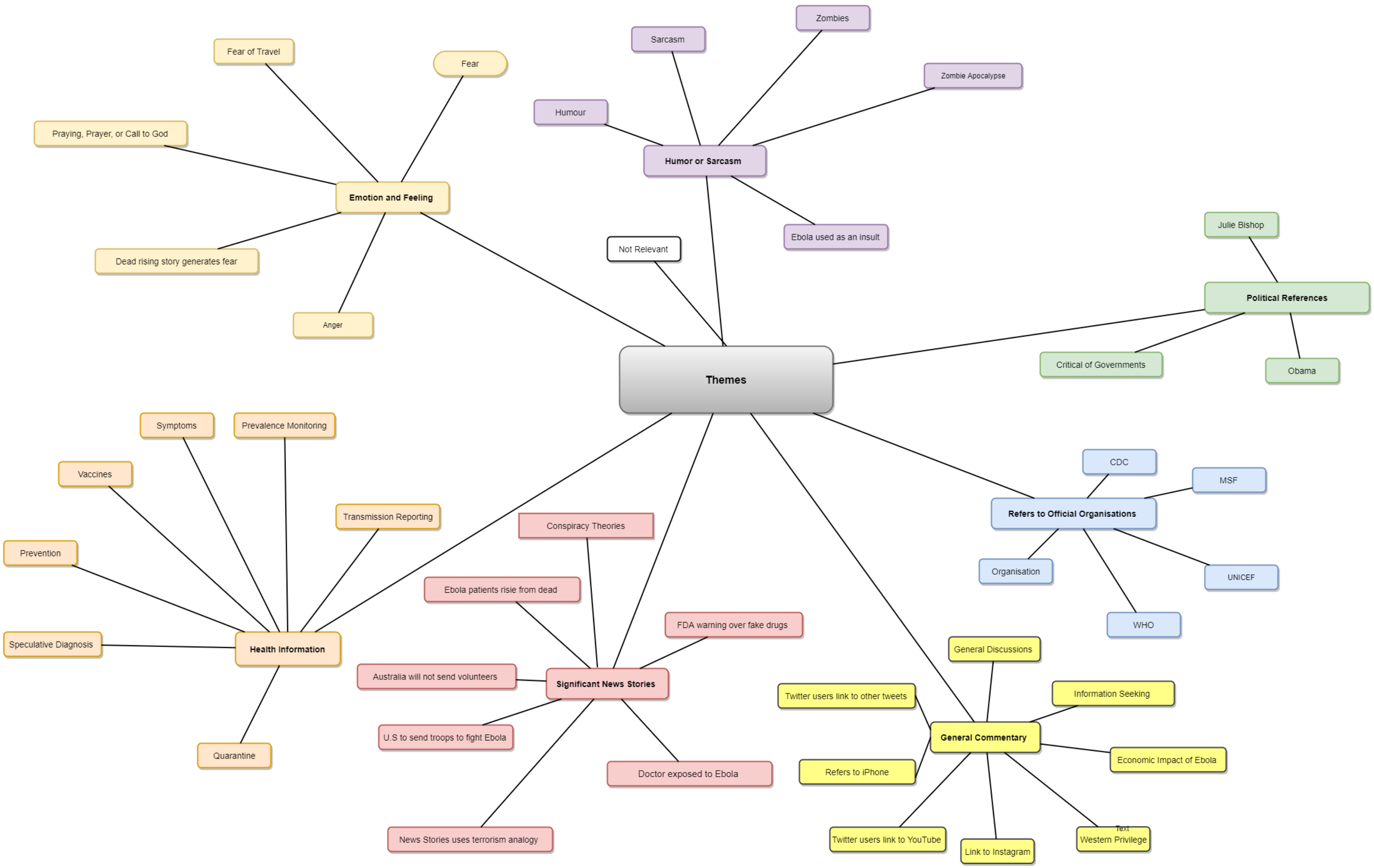
A total of 1401 tweets were discarded because they were either not in English (these were in Spanish), or in which content was purely spam based. The themes are not labelled from A, and instead begin with E to facilitate comparison to the swine flu case study in Chapter 7 Discussion.

Table 5- 3 Overview of themes and sub-themes (Ebola)

Theme (N/%)	Sub-themes (N/%)
E. Emotion and Feeling (113/2.60%)	E.1 Anger (12/0.3%) E.2 Fear (55/1.3%) E.3 Fear of travel (5/0.11%) E.4 Praying, Prayer or call to God (26/0.60%) E.5 Dead rising generates fear (15/0.35%)
F. Health Information (192/4.5%)	F.1 Transmission Reporting (41/1.00%) F.2 Transmission of Ebola (26/0.60%) F.3 Symptoms (37/0.90%) F.4 Vaccines (36/0.80%) F.5 Prevention (22/0.51%) F.6 Speculative Diagnosis (7/0.16%) F.7 Quarantine (25/0.60%)
G. Significant News Stories (282/6.60%)	G.1 Ebola Patients Rise from Dead (107/2.5%) G.2 Australia will not Send Volunteers (64/1.5%) G.3 US to Send Troops to Fight Ebola (18/0.42%) G.4 News Story uses Terrorism Analogy (6/0.14%) G.5 Doctor Exposed to Ebola (87/2.03%) G.6 FDA Warning Over Fake Drugs (30/0.70%)
H. General Commentary (2311/54.0 %)	H.1 General Discussions (2025/47.26%) H.2 Information Seeking (28/0.65%) H.3 Economic Impact of Ebola (11/0.25%) H.4 Death Count (32/0.74%) H.5 Western Privilege (11/0.25%) H.6 Link to Instagram (24/0.56%) H.7 Twitter Users Linking to YouTube (56/1.30%) H.8 Refers to iPhone (9/0.21%) H.9 Twitter Users Linking to Other Tweets (55/1.30%) H.10 Downplaying Ebola risk (9/0.21%) H.11 Conspiracy Theories (51/1.20%)
I. Refers to official organisations (75/1.80%)	I.8 WHO (17/0.40 %) I.9 CDC (37/0.90%) I.11 MSF (12/0.30 %) I.12 UNICEF (9/0.21%)
J. Refers to West African City and or region (181/4.2)	I. 1 Sierra Leone (104/2.42%) I.2 Liberia (36/0.84%) I.3 Nigeria (33/0.80%) I.4 Guinea (8/0.2%)
K. Political References (88/2.05%)	I. 1 Obama (68/1.58%) I.2 Julie Bishop (7/0.16%) I.3 Critical of Governments (13/0.30%)
L. Humour or Sarcasm (1046/24.41%)	J.1 Sarcasm (425/9.9%) J.2 Humour (418/9.8%) J.3 Zombies (77/1.8%) J.4 Zombie Apocalypse (18/0.42%) J.5 Ebola Used as an Insult (108/2.52%)

The figure below is a diagrammatical overview of the themes and sub-themes that emerged from the in-depth analysis of tweets on the Ebola epidemic.

Figure 5-2 Diagrammatical overview of themes and sub-themes (Ebola)



E. Emotion and Feeling

This theme contained tweets which expressed an emotion towards the Ebola outbreak and included sub-themes of anger, fear, fear of travel, prayer, and a specific news story which caused Twitter users to express fear.

Table 5-4 Sub-themes of emotion and feeling

Theme	Sub-themes
E.Emotion and Feeling	E.1 Anger
	E.2 Fear
	E.3 Fear of travel
	E.4 Praying, Prayer or calls to God
	E.5 Fear of dead rising

E.1 Anger

Some Twitter users came across as angry within their tweets and would mention the term Ebola. In certain instances, the anger would be expressed towards the Ebola outbreak. Illustrations of these tweets are provided below:

'The people in my area have Ebola, WHAT THE HELL!' (E1.1)

*'F**KING HELL with the Ebola news!' (E1.2)*

On the Internet, using text which is capitalised is often associated with shouting and/or the emotion of anger, which is a feature of the two tweets above. Additionally, there were also users who expressed anger in their tweets, but not necessarily towards the Ebola outbreak, as illustrated in the examples below:

*'F**KING user who just unfollowed me, hope you die of EBOLA' (E1.3)*

'Unable to play football online, servers are EBOLA' (E1.4)

It appears that the word Ebola had entered the vocabulary of Twitter posts during this time, and users were tweeting using the word Ebola for irony, rather than commenting on the outbreak. Themes were not mutually exclusive and tweets in this category of emotion and feeling may also have been coded as conveying sarcasm or humour. As highlighted in Chapter 5, the vocabulary adopted by Twitter users may seem to be exaggerated and this could be due to the age group of Twitter users which has been reported to be between mainly 18 to 24 year olds (Pew Research Center, 2016).

E.2 Fear

Some Twitter users expressed fear towards Ebola, and towards specific news stories which were published at the time. There were also some users fearful of Ebola entering their country, and of the possibility of Ebola mutating and infecting them. In the context of the Health Belief Model, as the perceived severity of Ebola is high, Twitter users may have been more afraid of Ebola than other viruses and diseases, such as the regular influenza virus. This is because Ebola is perceived to be more deadly than other viruses and diseases. Moreover, the symptoms of Ebola appear to be very unpleasant and have the potential of causing fear among the general public. Illustrations of tweets which conveyed fear are illustrated below:

'Wow, Ebola is scary' (E2.1)

'The death count of Ebola is scaring me' (E2.2)

'If Ebola comes to the U.S. I think we are finished, so scared' (E2.3)

'Very fearful of Ebola mutating, think we should help the affected' (E2.4)

The risk of a widespread Ebola pandemic in 2014 was relatively low as Ebola is not airborne (Centre of Disease Control, 2015). However, a number of Twitter users were afraid that the disease would spread rapidly across the US, although this was not likely to occur (Centre of Disease Control, 2015). A potential reason for the fear among users around this time could relate to the severity of Ebola. As the symptoms of Ebola are very frightening, the slightest risk of developing it may cause users to express fear. Moreover, it can also be argued that, as the symptoms of Ebola are out of the ordinary, users may be more afraid of the disease. Therefore, Ebola had the potential to generate negative feelings among the general public due to its perceived severity. Other users replied to news organisations by writing that they were afraid of the Ebola outbreak, for example, some Twitter users replied to a CNN story suggesting that the story they had run was scary, as illustrated below:

'@CNN That Ebola is scary!' (E2.5)

'@NBC Omg so scary reading that!' (E2.6)

It appears that news stories reporting on the Ebola epidemic increased fear of the disease. For example, in tweets *E2.5* and *E2.6,* users were replying to news stories indicating that what they had read had scared them. Therefore, the mainstream media could have played a role in perpetuating fear among the general public due to the types of articles that were being shared. For example, a story which noted that Ebola victims had risen from the dead was shared by a number of media outlets. Further research could seek to analyse news that was disseminated via the media such as online news stories to ascertain whether material was sensationalised and exaggerated by the media in an attempt to attract more readership.

Furthermore, there were tweets where it was difficult to know whether users were exaggerating the views they were expressing, or they were genuinely afraid. Illustrations of these tweets are provided below:

'I am sure I have Ebola, and that I will be dead soon' (E2.7)

'Ebola is so scary – people are crying!' (E2.8)

'Omg! If Ebola is reported in the U.S. I will move to Canada.' (E2.9)

It could be argued that users were experiencing some level of fear, but that the vocabulary they were using was exaggerated. As highlighted in Chapter 5, people online may post and behave differently to how they would in real life. A limitation of qualitatively analysing tweets is that it is difficult to interpret whether users are being honest or satirical. However, alternative methods such as sentiment analysis and/or machine learning approaches, which are automated, may misclassify tweets inaccurately if the tweets involve sarcasm.

E.3 Fear of travel

A number of Twitter users referenced how they felt afraid of travelling because of the Ebola outbreak. There were also Twitter users who were considering travelling to the African continent but altered their plans and did not travel:

'We thought of visiting Morocco, but are unsure now if it is safe for the children, don't want to get Ebola!' (E3.1)

'My mum is apprehensive of me travelling to Africa due to Ebola' (E3.2)

There were other users who noted that a family member had expressed travel related concerns and/or wanted to alter their plans due to the Ebola outbreak, as illustrated below:

'Grandmother doesn't want to travel to Morocco due to the Ebola outbreak' (E3.3)

'Father suggests that I not go to Asia because of Ebola and also a terrorist threat' (E3.4)

In the context of the Health Belief Model, as the perceived severity and seriousness of Ebola is high, users may have attempted to avoid a possible negative health outcome (contracting Ebola) by altering their plans to visit Ebola affected regions.

E.4 Praying, Prayer, or call to God

There were Twitter users who had been praying for family members who were in Ebola affected areas such as West Africa, or for the victims of Ebola, and hoped that God would be listening to their prayers:

'I pray for family who are in Liberia every single day' (E4.1)

'Will be praying for all those affected by the Ebola outbreak' (E4.2)

'Jesus can heal the Ebola patients – pray to Jesus' (E4.3)

'Anybody who is interested in a group praying session – let me know' (E4.4)

'All we need is God, and Ebola will be stopped' (E4.5)

'Hope that God is listening and acts!' (E4.6)

In the context of the Health Belief Model, being infected with the Ebola is a negative health outcome with deadly consequences. Therefore, Twitter users with relatives in West Africa would have had a very high perceived severity and may have taken up praying with the hope that their loved ones would be less likely to be in contact with Ebola. There is some scientific evidence suggesting that praying can reduce anxiety (Çoruh, Hana, Meredith and Thomas, 2005). Therefore, the Twitter users that had taken up praying during this time may have benefited from it.

E.5 Fear of the dead rising

A news story of deceased Ebola sufferers rising from the dead in Liberia generated fear and shock among the general public, and users hoped that the story was not factual. The story was not reported by mainstream media; however, a number of media outlets published it. In the story, there was the image of a 'zombie' taken and edited from the film *World War Z,* and the news story indicated that people had seen the person die, but that he later rose from the dead and regained life (Eleftheriou-Smith, 2014). Twitter users reading the news would react in shock and would note how the news story made them feel afraid. Illustrations of these tweets are provided below:

*'Ebola victims rise from dead? What the f**k?' (E5.1)*

'Wow, Ebola Zombies are actually real? That has to be the scariest thing I have ever read!' (E5.2)

'The story of Ebola victims rising from the dead better be a joke! Otherwise I will cry out of fear' (E5.3)

As this news story reported on an event that was out of the ordinary, it had the potential to cause fear among Twitter users. However, there is also the possibility that Twitter users may have been attempting to perpetuate fear and cause controversy rather than being genuinely afraid. The initial users' reactions could have been of shock and disbelief and, after reflecting on the story, they may have altered their views towards it.

F. Health Related Information

Tweets were coded in this theme and branched into the following sub-themes: health information and medical information such as transmission reporting, transmission, symptoms, vaccines, prevention, users who self-diagnose themselves with Ebola.

Table 5-5 Sub-themes of health-related information

Theme	Sub-themes
E.Health information	F.1 Prevalence Monitoring
	F.2 Transmission of Ebola
	F.3 Symptoms
	F.4 Vaccines
	F.5 Prevention
	F.6 Speculative Diagnosis
	F.7 Quarantine

F.1 Prevalence Monitoring

There were Twitter users who reported on the transmission of Ebola, for example, Twitter users would track the spread of Ebola in their tweets and would refer to locations of where diagnosed Ebola cases had occurred. It appeared that mapping the geographical spread of Ebola was popular during this time period:

'There is Ebola patient in DC? It is getting close' (F1.1)

'Ebola now in Bethesda' (F1.2)

'Ebola now in a fifth country' (F1.3)

In the context of the Health Belief Model, people may use information derived from tracking Ebola when considering whether their perceived susceptibility to Ebola is high or low. It appeared that, as diagnosed cases were reported in closer proximity to Twitter users, their perceived susceptibility and consequent fear of Ebola increased.

In Chapter 4 the term *'Citizen Disease Surveillance'* was developed in order to refer to this activity of monitoring the spread of an infectious diseases, and can also be applied to this sub-theme. Citizen disease surveillance was defined in this present study as: "the tracking, monitoring, and/or surveillance of a disease outbreak using open tools and methods" by members of the public.

F.2 Transmission of Ebola

Some Twitter users would reference means of transmission of Ebola, for example, how bush meat, monkeys and mosquitos could transmit the disease, while water could not:

'When Ebola is gone from the body, blood is no longer infectious' (F2.1)

'They have been eating bush meat for the last few years' (F2.2)

'To the best of my knowledge water does not spread Ebola' (F2.3)

There were also a number of Twitter users who would refer to elements of transmission:

'I think they caught Ebola from eating a monkey' (F2.4)

'I got bitten by mosquito, and now may have Ebola' *(F2.5)*

The illustrations above could also be intended as sarcastic and humorous but may reveal beliefs and associations that users may hold. For example, although the users may not be completely serious in their tweets above, a part of them may believe that monkeys or mosquitos can transmit the Ebola virus.

F.3 Symptoms

A number of Twitter users referred to some physical symptoms they were experiencing and equated this with the onset of Ebola. The symptoms ranged from throat pain, sudden fever, headache, and nose bleeds. Users outside of West Africa who suffered from symptoms of flu, fever or headache would express fear over their symptoms, and they hoped that they had not developed Ebola. It appeared that users who were tweeting from the West African region felt more afraid that they were exhibiting symptoms of the disease. In the context of the Health Belief Model, the concept of individuals feeling more afraid if they felt in closer proximity to Ebola is understandable, as their perceived susceptibility to Ebola was higher. Examples from this theme are provide below:

'My throat is hurting; I think I have Ebola' (F3.1)

'Sudden fever, I swear knowing my luck it will be Ebola' (F3.2)

'When you have a headache don't you just think you have Ebola? I do' (F3.3)

'I Googled my symptoms as I have sore throat and ear, WebMD said I may have Ebola.' (F3.4)

'Just had a nose bleed, definitely Ebola.' (F3.5)

'Living in Africa and getting sick makes automatically think it is Ebola' (F3.6)

Due to increased media coverage during the outbreak, when Twitter users experienced symptoms such as headache, fever, or a sore throat, they may have thought of Ebola first. A Twitter user noted that their flu symptoms may worry staff at an airport:

'My flu symptoms will make airport people think I have Ebola, down with fever too!' (F3.7)

This is an understandable anxiety as airports were screening passengers for potential high temperatures (Jansen, 2014). There were also users who referred to symptoms in other people, for instance, a user was concerned with coming into contact with members of the public who were sneezing. There were some Twitter users who recalled accounts of their being in close contact with people who were experiencing symptoms such as coughing or sneezing:

'So many people sneeze around me, I can sense Ebola in the air' (F3.8)

'This guy in Costa is coughing what looks like Ebola particles' (F3.9)

The Health Belief Model suggests that the likelihood of someone changing their behaviour would be influenced by the threat of a disease. The likelihood of a behavioural change would depend on factors such as severity and perceived susceptibility. Ebola is a fatal disease and has a high perceived severity. Therefore, users in the tweets above may have felt uncomfortable coming across members of the general public who were coughing and sneezing. This also indicates that there was sensitivity around symptoms that could potentially spread Ebola. This may be related to the increased media coverage during the outbreak.

F.4 Vaccines

There were users who referred to the lack of an effective vaccine for Ebola and how this would affect the developing world. There were also Twitter users who noted that American citizens were receiving vaccines over people from West Africa. Other Twitter users questioned whether anti-vaccination parents would refuse the Ebola vaccine. Additionally, news articles referring to warnings over fake FDA drugs were shared:

'With current mortality rate & without vaccine – Ebola may become a serious problem in developing world by end of year' (F4.1)

'Why is there no vaccine for Ebola yet?' (F4.2)

'So two Americans are given Ebola vaccine when so many Africans have died' (F4.3)

'I wonder if those anti-vaccination parents would actually refuse an Ebola vaccine for their child if it spread to U.S?' (F4.4)

'FDA warns companies about fake Ebola drugs' (F4.5)

There have been a number of discussions around vaccines. In tweet *F4.1,* a user is referring to the danger of not having a vaccine as this would result in a higher mortality rate. In example *F4.2* a Twitter user is seeking to know why there is no vaccine for Ebola. Twitter accounts that were responsible for informing the public could have disseminated information on vaccines in order to address the information needs of users during the outbreak of Ebola, as this information was lacking. In tweet *F4.3,* users noted that Americans had been provided with vaccines whereas many Africans had died without one. There were a number of tweets with variations of this aspect. These types of tweets suggest that users felt there was some form of prejudice taking place in how vaccines were developed and distributed. Health authorities in the United States may wish to better inform the general public on how vaccines are allocated. Tweet *F4.5* uses the term 'anti-vaccination' in reference to parents who refuse to vaccinate their children to protect them from potential negative side-effects of the vaccines. The user seeks to know whether parents who are against vaccinations would refuse Ebola vaccination treatments for their children.

F.5 Prevention

Twitter users would refer to a number of prevention behaviours that could be taken in order to contain Ebola such as regularly washing hands, having a good level of personal hygiene, and wearing facemasks:

'Wash your hands or you can spread Ebola' (F5.1)

'Ebola has made me think more about personal hygiene' (F5.2)

'When I heard about the Ebola outbreak, it made me want to wear a mask outdoors' (F5.3)

Many of the prevention behaviours mentioned by users were also those advised by health authorities.

F.6 Speculative Diagnosis

A number of Twitter users would speculatively diagnose themselves with Ebola. However, users may have tweeted that they have Ebola as a joke:

'What if it turns out I have Ebola, and CDC just ignores me!' (F6.1)

'John is coughing and now he thinks he has Ebola' (F6.2)

'I have Ebola… self-diagnosed so might not be true' (F6.3)

The author of tweet *F6.1* had speculatively diagnosed themselves with Ebola, and provoked the CDC's Twitter account. However, it is most likely that the user had not developed Ebola. This highlights the difficulty of taking tweets at face value, as Twitter users may be playing up to their followers and attempting to create controversy among their peers.

F.7 Quarantine

Several Twitter users referred to the concept of quarantine:

'Without effective quarantine procedures Ebola will become rampant across the developing world' (F7.1)

'Troops that return from Ebola affected areas should be quarantined' (F7.2)

'So scary that the Chief Doctor in Liberia is in quarantine' (F7.3)

In tweet *F7.1,* the user is referring to potential negative implications that a lack of quarantine would have among the general public in the developing world. This is because, without effective quarantine in place, it may be difficult to halt the propagation of a disease. Tweet *F7.2* is referring to the idea that people who enter Ebola affected areas should be quarantined upon returning. Referring to the Health Belief Model, this type of reasoning can be thought of as aiming to mitigate the potential threat of Ebola to an individual and a community as a whole. This is because, as the perceived severity of Ebola is high, users may have thought of the potential negative impact that people returning from Ebola affected areas would have on their health and well-being. There were also Twitter users who would refer to a Liberian chief doctor who had been quarantined during this time period.

G. Significant News Stories

This theme contained a series of sub-themes (Table 5-6), which referred to a specific news story that was significant at the time of the outbreaks. There were a number of news stories shared on Twitter during the outbreak. However, the news stories coded within this sub-theme appear to have been particularly significant at this time as they were either shared frequently, commented on or referenced in tweets, as shown in Table 5-17.

Table 5-6 Sub-themes of significant news stories

Theme	Sub-themes
G. Significant News Stories	G.1. Ebola patients rise from dead
	G.2 Australia will not send volunteers
	G.3 US to send troops to fight Ebola
	G.4 News story uses terrorism analogy
	G.5 Doctor exposed to Ebola
	G.6 FDA warning over fake drugs

G.1 Ebola patients rise from dead

A variant of this sub-theme was reported in theme E, but related more to the fear that the story generated. In addition to users replying on how they felt after reading the story, they also tweeted it out directly and modified slightly in order to express an opinion within a tweet. Illustrations of tweets that shared this story are provided below:

'Ebola victims rise from dead in an African village' (G1.1)

'Panic in Liberian Village as Ebola Victims Rise from Dead' (G1.2)

'Apparently Ebola victims rose from the dead – can anyone actually verify this story?' (G1.3)

*'Heard the news of Ebola victims rising from dead – *preparing for Zombie apocalypse*' (G1.4)*

'Hope story of Ebola patients rising from dead is a joke – I am about to cry out of fear' (G1.5)

The story had no factual basis (Eleftheriou-Smith, 2014) and was not reported by the mainstream media, and a number of Twitter users questioned the validity of the story. The story could be considered poor journalism. The sensationalist and 'out-of-the-ordinary' headline may have attracted high page views and attention on social media.

Some Twitter users questioned the idea that people believed the news that Ebola victims had come back to life. For example, a user made fun of people who would take the story at face value, these examples are illustrated below:

'LOL, people actually believe Ebola victims came back to life' (G1.6)

'Don't understand this story of Ebola bringing people back to life – how can this be true?' (G1.7)

This suggests that Twitter users did not necessarily believe the story, and that at least a sub-set of them thought the story was not factual. Moreover, users may have known that the story was false, but wished to share it with their peers for dramatic effect and to create controversy. There may be users who lack a formal education, and could have believed the story. These findings have a number of implications for health authorities such as that they could provide up to date information when rumours such as this would arise and this will be further explored in the discussion of results. Within the sample of tweets analysed as part of this project, no tweets or news articles from official health bodies were identified that were sent in order to contradict the story of Ebola victims rising from the dead.

G.2 Australia will not send volunteers

A news story shared by a number of users made reference to the Health Minister of Australia who stated that Australia would not send health workers to West Africa because there were no mechanisms in place to bring them back safely. Twitter users were critical of Australia's stance, and some suggested that it was an issue of race. Other users noted that Australia had previously supported the war in Iraq by sending military personnel, and that it seemed unfair to not send support to West Africa. Illustrations of these of tweets are provided below:

'Ebola outbreak: Australia will not send health workers to west Africa' (G2.1)

'Not sure why Australia reluctant to send health workers to fight Ebola, but they're quite keen going to war' (G2.2)

'Australia are not sending help to fight Ebola' (G2.3)

'Apologies to people who are brown- looks as Australia does not do non-military humanitarian assistance' (G2.4)

There were tweets which re-shared the news story that Australia would not send health workers to Ebola affected areas without offering a personal opinion about it. However, there were Twitter users who expressed the view that Australia had purposely avoided sending health workers to fight Ebola, and accused the government of prejudice. Moreover, some users suggested that people of colour were being discriminated against. The official reason provided by the Australian government for not sending troops to Ebola affected areas related to logistics. The government claimed that if health workers were to be infected with Ebola, then there would be no safe way to bring them back. Twitter users noted that the Australian government had sent troops to war in Iraq, but was not sending health workers. However, this may be an unfair comparison. When troops are sent to fight war, there are risks of a different nature to those of a large-scale infectious disease outbreak.

G.3 US to send troops to fight Ebola

A number of Twitter users reported the news that the US had sent troops to West Africa in order to combat Ebola, and there were also users who would offer an opinion towards the story:

'US is sending 3000 troops to Liberia to combat Ebola' (G3.1)

'U.S. Troops Battling Ebola get off to a Slow Start in Africa' (G3.2)

'They're sending troops to Liberia to fight Ebola or is it to kill people with Ebola? How weird to send troops for a disease' (G3.3)

'Hurray to us! We are happy to send military to war zones' (G3.4)

'US troops in Liberia to fight Ebola? So when will plane load of health workers arrive?' (G3.5)

During an infectious disease outbreak, governments may send health workers to the affected geographical areas in order to help mitigate its spread. Twitter users may have been surprised to find out that troops rather than health workers were sent to West Africa to help with the relief effort. Troops are normally sent when there is a conflict or a war, or following a natural disaster such as earthquakes. Therefore, it is not surprising to see tweets such as *G3.3, G3.4* and *G3.5* which either questioned or mocked the decision of the United States to send troops to Ebola affected areas.

G.4 News Story uses Terrorism Analogy

A news story which was tweeted out by a number of users suggested that Nigeria would have defeated terrorism if the same effort that was applied to fighting Ebola was applied in fighting it. This article was shared by a number of Twitter users. Other news articles which employed a terrorism analogy suggested that the current outbreak of Ebola was worse than terrorism. Illustrations of these of tweets are provided below:

'Nigeria would have conquered Terrorism if only we fought against it like we fought Ebola [URL]' (G4.1)

'Ebola is worse than terrorism – according to Sierra Leone Leader [URL]' (G4.2)

The types of news stories above did not incite a response from Twitter users, and were simply tweeted out and shared. Although Ebola is a different type of threat to terrorism, journalists may have seen parallels with the regional fight against terrorism. Users would either tweet this story or retweet a tweet about the story without offering an opinion. It is difficult to ascertain the motives of the users who were sharing these stories, for example, whether the users who tweeted this story agreed with the views expressed in them, found them to be interesting, or simply wished to share information with peers. Research into retweeting has found that people might retweet for a number of reasons. These could include a user publically agreeing with someone else, and to validate another user's views (boyd, Golder, and Lotan 2010).

G.5 Doctor Exposed to Ebola

There were Twitter users who were sharing information related to a US doctor who had been exposed to Ebola:

'US Doctor Exposed to Ebola Virus Admitted to NIH' (G5.1)

*'This Fu**ing doctor should not be brought back to Maryland, take him to another continent' (G5.2)*

'The doctor and all others in the same situation should be banned from re-entering the U.S. – they are aware of the risk they take' (G5.3)

The doctor had been volunteering in Sierra Leone and was diagnosed with Ebola. In tweet *G5.1*, a user referred to the National Institute of Allergy and Infectious Diseases (NIH). There were Twitter users who thought that the doctor should not be brought back to the US, presumably because there would be a chance that the virus could spread across the population. For example, in tweet *G5.2* the user noted that a doctor who was exposed to Ebola should be taken elsewhere and not transported back to the US. Similarly, in tweet *G5.3*, a Twitter user expressed the view that all doctors who accidently contracted Ebola when volunteering should have been banned from entering the US. The user may have expressed this view as they may have been thinking of their own risk and the potential risk of an epidemic to the US. Stigma towards health workers can be harmful as it could reduce the number of volunteers, as well as potentially make them fearful of coming forward if they were to be infected by Ebola.

G.6 Liberian Health Minister

There was a number of Twitter users sharing a news story of a Liberian Health minister who had voluntarily entered quarantine after their office assistant had died of Ebola:

'Liberian Health Minster Quarantines Self after Office Assistant Died of Ebola [URL]' (G6.1)

'Saw the story of the Health minister under quarantine – look at the damage it has caused' (G6.2)

'Ebola Liberia health chief quarantine demonstrated challenge of Ebola as providers of services

are also affected' (G6.3)

This story may have been particularly newsworthy as it should be expected that those working within the government health departments would have taken precautions against Ebola. Tweet *G6.2* noted that this particular story highlighted the extent of the damage that Ebola had caused.

G.7 FDA Warning Over Fake Drugs

Some Twitter users were sharing an article from the United States Food and Drug Administration (FDA). The FDA had sent a warning letter to a number of companies who had been falsely claiming that their drugs could cure Ebola:

'The FDA offer warnings over Fake Ebola drugs by issuing letters' (G7.1)

This is an important article, and Twitter users may have wished to share this information with others to warn them of potential fake drugs on the market. The organisations marketing bogus Ebola drugs may have thought that the Ebola epidemic was something they could profit from. From a public health perspective, Twitter may have acted as a valuable mechanism for disseminating warnings to the general public on the danger of fake drugs.

H. General Commentary

There was a number of Twitter users who would make general comments towards the Ebola outbreak, and these tweets were coded into this theme.

Table 5-7 Sub-themes of general commentary

Theme	Sub-themes
H. General Commentary	H.1 General commentary and/or news
	H.2 Information seeking
	H.3 Economic impact of swine flu
	H.4 Death count
	H.5 Western privilege
	H.6 Link to Instagram
	H.7 Twitter users linking to YouTube
	H.8 Refers to iPhone
	H.9 Twitter users link to other tweets
	H.10 Downplaying Ebola risk

H.1 General Commentary and/or News

In some tweets, users would make general comments and statements about Ebola. There was also a wide range of news articles that may have been shared. Some tweets were difficult to interpret as they did not express an idea or a view that could be coded. Illustrations of these tweets are provided below:

'World is still acting slowly to Ebola [URL]' (H1.1)

'Just found out about the Ebola epidemic' (H1.2)

'People keep talking about Ebola' (H1.3)

'This is an interesting story on Ebola [URL]' (H1.4)

This sub-theme was the largest for the number of tweets that were coded in it. In an interview or a survey, participants' responses can be directed and focused towards a topic under study, whereas on Twitter there may be many comments and/or news articles shared which lack depth. Previous evidence-based research which examined swine flu (Chew and Eysenbach, 2010) similarly found that the largest categories of tweets were those which were resource based, i.e. which provided links to news or general updates.

H.2 Information Seeking

There were some users who were seeking information in their tweets:

'What is Ebola?' (H2.1)

'What have the WHO been doing about Ebola' (H2.2)

'Is there Ebola in Ghana?' (H2.3)

'Is there really a person with Ebola in Washington DC?' (H2.4)

It is difficult to know whether some of the users were posing questions rhetorically, or they had genuine information needs. Twitter users may have been posting questions related to Ebola as they may have been exposed only to partial information. For example, in tweet *H2.4,* the user may have been exposed to the information that there was Ebola in Washington D.C. and may have wanted to get confirmed by other users. In tweet *H2.2* the user has been asking what the WHO had been doing to tackle Ebola as this was what they would be expected to be doing.

H.3 Economic Impact of Ebola

Several Twitter users were referring to the economic impact of Ebola within their tweets:

'The price of cocoa has fallen dramatically because of Ebola [URL]' (H3.1)

'Ebola has taken a big toll on Kenyan tourism [URL]' (H3.2)

'There have been calls for people to stop Ebola from negatively affecting the African Economy' (H3.3)

'Maybe when people find out what has been happening to coca prices they will wake up about Ebola' (H3.4)

There were users who were tweeting out news stories which would refer to the economic impact of the Ebola outbreak, for example, tweets *H3.1* and *H3.2*. However, there were also users such as *H3.3* and *H3.4* who were tweeting their own views regarding the economic impact of Ebola.

H.4 Death Count

There were some Twitter users who were referring to the number of deaths that the outbreak of Ebola had caused:

'According to the CDC the death toll due to Ebola could hit 550' (H4.1)

'Death toll because of Ebola is now at 887' (H4.2)

'Death toll because of Ebola is now at 3000' (H4.3)

Interestingly, it appeared that the numbers cited by users within their tweets varied. This could be either because the sources users were referring to were outdated and consequently contained lower figures, or because users were measuring the rate of deaths in different geographical areas such as specific countries and continents. For example, in tweet *H4.1*, the Twitter user referred to the death toll specifically in West Africa. Only one death took place outside of West Africa during the epidemic, and that was in the US (BBC News, 2016).

H.5 Western Privilege

A number of Twitter users referred to those living in Western nations being more privileged than those born in developing countries. For example, people might perceive that those who live in developed countries, such as the US, would receive vaccines faster than people who were from developing countries, such as Guinea. Examples of some of these tweets are illustrated below:

'No one cares about Ebola because of racism' (H5.1)

'I am so upset that a lot of people are dying every day due to Ebola in West Africa because of Ebola but no one cares!' (H5.2)

'Why did two Americans get Ebola serum whilst hundreds of people in West Africa died?' (H5.3)

'Ebola is killing so many people in Africa and the news interview someone white!' (H5.4)

There are different views expressed by Twitter users in the messages above. Users may have felt that people were not concerned with the Ebola outbreak, and that race may have played a role in the fight against Ebola. An expert panel in a report published in the Lancet found that the response to the outbreak was too slow, and that health authorities should have acted faster (Moon *et al.*, 2015). Therefore, at the time of the outbreak, users could have picked up on the fact that the response was slow and inadequate and tweeted about this. In addition, those in developed countries were more privileged than those in developing ones, especially with regards to how vaccines were distributed. This view was also expressed by online writers during that time (Williamson, 2014), especially in terms of the lack of transparency of vaccine distribution.

H.6 Link to Instagram

Some Twitter users were linking to the image-sharing platform Instagram within their tweets. Examples of these tweets are illustrated below:

'Who said zombies were just to be found in video games? [Link to Instagram]' (H6.1)

'Nigeria is still Ebola free [Link to Instagram]' (H6.2)

Twitter users would link Ebola-related images from Instagram to Twitter. However, there were Twitter users who would link to Instagram, but the Instagram post would not contain any images related to the Ebola outbreak. This is known as 'link-baiting'. It is important to note that as Instagram and Twitter are two different commercial entities, it is not possible for Instagram posts to natively appear on Twitter, as they can only be linked to it.

H.7 Twitter users linking to YouTube

A number of Twitter users were linking to the video-sharing platform YouTube within their tweets as illustrated in the examples below:

'I liked a YouTube video on what it is like to live in a country affected by Ebola [Link to YouTube video]' (H7.1)

'Doctor who was exposed to Ebola is now under surveillance [Link to YouTube video]' (H7.2)

Videos from YouTube can often be played within tweets and without having to navigate away from the Twitter platform. Twitter users would link to informational videos from YouTube and share these on Twitter. The videos shared tended to relate to news items of the outbreak and were commonly viewed on the platform at that time. Videos were not analysed as part of this present study.

H.8 Refers to iPhone

There was a number of Twitter users who were referring to how the 'iPhone 6' had been contaminated with Ebola:

'Everyone is talking about iPhone being contaminated with Ebola – wish people would look at the sources they refer to' (H8.1)

'I refused an iPhone 6 today because of the Ebola outbreak' (H8.2)

'2000 iPhone 6 devices were contaminated with Ebola' (H8.3)

'On my Facebook it said up to 60% of iPhones are affected by Ebola' (H8.4)

Previous evidence-based research looking at rumours on Twitter (Jin *et al.*, 2014) provided similar findings. This is because in the study by Jin *et al.* (2014) it was noted that the iPhone 6 was infecting people with Ebola. The findings in the present study show that users would refer to either the iPhone or iPhone 6 in relation to the devices being contaminated with Ebola. From the tweets themselves, it is not known to what extent (if any) Twitter users genuinely believed that the iPhone 6 had been contaminated with Ebola or whether users were sharing this information in order to create controversy and/or to target rival phone types.

H.9 Twitter users linking to other tweets

There was a number of users who were linking to other tweets within their own tweets. Users would do this either by linking directly to the URL or by quoting the tweet:

'Why would we doubt this? [link to quoted tweet]' (H9.1)

'These are the symptoms of Ebola [link to tweet]' (H9.2)

Some Twitter users may have quoted others as a way of providing context before commenting on a tweet. This practice suggests that tweets may have been an information source in their own right.

H.10 Downplaying Ebola Risk

Some Twitter users were downplaying the risk of an outbreak of Ebola in the West within their tweets:

‘There is not really a chance of an Ebola outbreak in America’ (H10.1)

‘All of the true experts know that Ebola is not a big deal. It is just an issue for third world countries’ (H10.2)

Twitter users also held the belief that Ebola would not affect developed countries, a view expressed in tweet *H10.2.*

H.11 Conspiracy Theories

A number of Twitter users shared conspiracy theories related to the Ebola outbreak. People were sharing news of a professor at a US university who was suggesting, in a Liberian newspaper, that Ebola was manufactured by the US government:

‘Major publication in Liberia speculates that Ebola is a funded US bioweapon’ (H11.1)

‘Ebola just comes out of nowhere and it is out of control, it can’t be an accident that it is in West Africa where most black people live’ (H11.2)

‘Did the U.S create Ebola? As a form of population control?’ (H11.3)

‘There are rumours that Ebola is a big trick by doctors so they can steal blood from people’ (H11.4)

‘Ebola will not spread here – as they purposely poisoned people’ (H11.5)

‘Ebola is a massive pharmaceutical fraud by the west’ (H11.6)

‘*U.S Professor has been telling Liberians Ebola was caused by U.S government*’ *(H11.7)*

The tweets above express a range of different conspiracy theories shared by users. It is important to understand the stories behind these conspiracy theories, which suggested that infectious diseases are spread by the government. Between 1952 and 1972, the US government conducted a number of experiments on African Americans, known as the

'Tuskegee syphilis experiments', without consent (Goertzel, 1994). During the AIDs epidemic of the 1980s, it was believed that AIDs was a man-made virus created by the government (Goertzel, 1994). In the case of Ebola, tweets *H11.1, H11.2, H11.3* and *H11.5* seem to express the view that Ebola was manufactured by the US government.

In one of the few studies discussing conspiracy theories, Goertzel (1994) conducted a survey on 348 residents in south-western New Jersey. The author found that respondents thought a list of ten commonly believed conspiracy theories were most likely to be true. He also found that respondents from Black and Hispanic backgrounds were more likely to believe in conspiracy theories. Goertzel concluded by noting that conspiracy theories were found to exist in contemporary American society. This could explain why a number of conspiracy theories emerged on Twitter.

Storr (2014), in a Telegraph article (a British Broadsheet Newspaper) published during the 2014 epidemic, reported on some of the commonly held conspiracy theories at the time. The first of these was that Ebola was a bio-weapon developed by the military that had been accidently released. This conspiracy theory was propagated by a professor in the US. This view was expressed in Tweet *H11.7*. The professor who was spreading the conspiracy theory was an expert in this field since he served on the government's Committee of Military Use of Biotechnology, and also authored the Biological Weapons Anti-Terrorism Act of 1989. Biological weapons were developed throughout World War I and World War II in the US as well as France, Japan and the UK (Guillemin, 2006). Based on events of the past and the previous involvement of the US in bio-weapons, some Twitter users viewed Ebola as a bio-weapon.

In additional to a number of tweets sharing conspiracy theories, there were users sharing news articles which noted that Ebola conspiracy theories were spreading in the region, as illustrated below:

'Ebola conspiracy theories spread in Liberia [URL]' (H11.8)

This suggests that the authors of the article above may have had reasons to believe Ebola had been spreading in Liberia. A Twitter user referred to users who share conspiracy theories in negative terms as illustrated below:

'You know what, conspiracy theorists support wacko ideas with very selective and cherry picked evidence' (H11.9)

The user in the tweet above may have been aware of the tweets on conspiracy theories, leading them to send the tweet in the first place.

I. Refers to Official Organisations

A number of Twitter users referred to official organisations such as the World Health Organisation (WHO) and the Centre For Disease Control (CDC). Charitable organisations were also tasked with the relief effort of Ebola such as Medecins Sans Frontieres (MSF), and The United Nations Children's Fund (UNICEF).

Table 5-8 Sub-themes of official organisations

Theme	Sub-themes
I. Refers to Official Organisations	I.1 Reference to the WHO
	I.2 Reference to the CDC
	I.3 Reference to the MSF
	I.4 Reference to UNICEF

I.1 Reference to the WHO

A number of Twitter users tweeted news from the World Health Organisation (WHO) and retweeted information provided by the WHO:

'WHO: Ebola is infectious after death so pay your respects from at least a metre away' (I1.1)

'World Health Organisation: Ebola has killed 932 citizens across 4 countries' (I1.2)

'WHO: A number of Ebola vaccines are due to be rolled out' (I1.3)

'Thanks to the delay by the WHO, if it continues like this then people infected with Ebola may rise to 1.4 million' (I1.4)

The WHO is expected to disseminate information and help mitigate and fight against infectious disease outbreaks. The tweets above refer to different aspects of the Ebola outbreak on which the WHO were commenting on.

I.2 Reference to the CDC

Some Twitter users referred to the news that the Centre for Disease Control (CDC) were disseminating at the time of the outbreak:

'Because of CDC estimations if there are enough Ebola centres in the next 6 months we can put a halt to this epidemic.' (I2.1)

'CDC estimates there could be up to 1.4 million people who have Ebola' (I2.2)

'The CDC are telling us to prepare for the Ebola epidemic' (I2.3)

'According to the CDC, 70% of people who are infected must receive isolation for the Ebola to decrease' (I2.4)

It appears that Twitter users perceived the CDC to be an authority figure, and tweeted and retweeted news stories disseminated by the organisation. Twitter users referred to a range of different news articles mentioning the CDC, ranging from estimates on incidence to the importance of isolating infected individuals.

I.3 Reference to MSF

A number of Twitter users were referring to Medecins Sans Frontieres (MSF) in their tweets:

'MSF president is to address a High-Level UN meeting on Ebola' (I3.1)

'A lot of the people who work in Ebola treatment centres were trained by the MSF' (I3.2)

'MSF: For $600 you can buy 200 tablets which can relieve painful Ebola symptoms. You have power to make a difference' (I3.3)

It appeared that when Twitter users refer to the MSF, they do so positively. The MSF is an international network made up of volunteers with the role to assist those who might be affected by epidemics and of other disasters. The MSF might have been viewed as more independent than government-funded organisations.

I.4 Reference to UNICEF

Some tweets were referring to The United Nations Children's Fund (UNICEF) within their tweets:

'UNICEF teaching people how to make their own protective gear: because of this a women saved 3 people' (I4.1)

'UNICEF: how does Ebola spread? Here a list of ways [URL]' (I4.2)

'@UNICEF: Praise to the people who are fighting Ebola' (I4.3)

'RT @UNICEF Look at the impact Ebola is having on education: start of school year postponed indefinitely in Sierra Leone [URL]' (I4.4)

UNICEF is a charity seeking to help children in danger, and which was actively supporting the relief efforts for Ebola in West Africa in 2014. UNICEF was also sending a number of informational tweets. For example, tweet *I4.2* contains an infographic that was retweeted widely by Twitter users. A news story popular around this time related to the organisation teaching people how to make their own protection gear against Ebola.

J. References to West African Cities or Region

A number of Twitter users referred to specific geographical areas and regions within West Africa, which was the origin of the Ebola outbreak. The purpose of this theme is to highlight the various regions that were mentioned during the outbreak.

Table 5-9 Sub-themes of references to West African cities or region

Theme	Sub-themes
J. References to West African Cities or Region	J.1 Reference to Sierra Leone
	J.2 Reference to Liberia
	J.3 Reference to Nigeria
	J.4 Reference to Guinea

J.1 Reference of Sierra Leone

Some Twitter users referred to events related to Sierra Leone within their tweet:

'US Doctor has been exposed to Ebola in Sierra Leone' (J1.1)

'NHS staff will join the fight against Ebola in Sierra Leone' (J1.2)

'Here are pictures from the fight against Ebola in Sierra Leone' (J1.3)

'As the sun will be setting over Ebola centres in Sierra Leone – really sad to think some people will not make it to the next day' (J1.4)

Sierra Leone was mentioned on Twitter during the outbreak for a number of reasons. One reason was that a doctor from the United States was exposed to Ebola here, which seems to have been particularly newsworthy at the time, as reported by tweet *J1.1*. In tweet *J1.2*, the user is referring to the National Health Service (NHS) from the UK as a number of staff was sent to help with the relief effort. Tweet *J1.3* refers to a number of pictures shared from Sierra Leone and related to the outbreak of Ebola.

J.2 References to Liberia

There were references to the West African country of Liberia from Twitter users within their tweets, and these tweets are illustrated below:

'They promised Ebola Aid and It has fallen short of what they needed in Liberia [URL]' (J2.1)

'Liberia is in chaos due to the Ebola outbreak' (J2.2)

'Doctor has managed to treat Ebola by using a HIV drug in Liberia' (J2.3)

'Liberian Health chief currently quarantined over Ebola' (J2.4)

The tweets above refer to different events and general news that were associated with Liberia. In tweet *J2.1*, the user refers to a lack of aid, whereas in tweet *J2.2*, the user tweeting is referring to the consequences of the Ebola outbreak. These findings indicate that Twitter users felt that Ebola had a negative effect on Liberia.

J.3 References to Nigeria

There were references to the West African country of Nigeria from Twitter users within their tweets, and examples of these are illustrated below:

'How Nigeria has been containing the spread of Ebola' (J3.1)

'Ideas for a birthday gift – get me a vaccination pack for my trip to Nigeria' (J3.2)

'Government of South Africa must now close borders to people who are from Nigeria' (J3.3)

'Nigeria would have defeated Ebola if it was fought like they fought terrorism' (J3.4)

News stories were referenced reporting how Nigeria had dealt with the outbreak situation (*J3.1*), and how this country would have defeated Ebola if it fought as hard against terrorism (tweet *J3.3*). Tweet *J3.2* refers to a user who would be travelling to Nigeria, expressed concern about Ebola and shared views in a humorous manner.

J.4 References to Guinea

A number of Twitter users were referring to the West African country of Guinea within their tweets:

'The Ebola crisis may have profound economic consequences: the IMF will donate $130 to Guinea' (J4.1)

'Fear of Ebola in Guinea drove a mob to kill officials' (J4.2)

Many mentions of Guinea were from news articles, for example about the amount of money that had been donated to the population and information about localised events in Guinea.

K. Political References

This theme included tweets which made reference to politics or a political figure and includes the sub-themes of President Obama, Australian Foreign minister Julie Bishop, and criticism of government.

Table 5-10 Sub-themes of political references

Theme	Sub-themes
K. Political References	K.1 Obama
	K.2 Julie Bishop
	K.3 Critical of governments

K.1 Obama

A number of news articles shared by Twitter users mentioned the US president at the time of the Ebola outbreak in 2014. The most popular article related to President Barack Obama stated that the Ebola outbreak was a US national security priority:

'President Obama: Ebola is a U.S national security priority [URL]' (K1.1)

'Obama calls for a global response to the Ebola outbreak [URL]' (K1.2)

Every time Obama issued a statement via the White House, this would then be reported across the media and shared on Twitter. The news articles related to Obama were mostly neutral, however, there were some tweets which had a critical tone:

'Our so-called president Obama still has not come out with a plan to address Ebola' (K1.3)

'Can we even trust Obama with national security, and the Ebola outbreak?' (K1.4)

'Obama does not send troops to fight Muslims, yet he will send troops to save Muslims from Ebola' (K1.5)

''Obama really sees soldiers as disposable – whether losing soldier in Afghanistan to IED or to Ebola is West Africa' (K1.6)

The users critical of Obama focused on different aspects of the outbreak situation, and blamed Obama for some of the events that occurred. In tweet *K1.3*, a user is critical of Obama due to the lack of a plan to fight Ebola. Tweet *K1.4* complains that Obama cannot be trusted with national security. These users may have already have felt resentment towards Obama and may have used the outbreak of Ebola as a political weapon. A number of Twitter users shared the view that Obama would not send troops to fight Muslims because he had an affinity with them. For example, in tweet *K1.5*, a user notes that Obama was reluctant to send troops to fight “his Muslim friends and family”. Political opponents would also state that he was a Muslim when Obama was running for president, as well as when he was in office as president (Bredderman, 2016). In tweet *K1.7*, a Twitter user criticises Obama for sending troops to war zones and to Ebola affected areas. The criticisms levelled at Obama seem inconsistent, since Obama was criticised whether he sent aid to West Africa or not. These results may be of interest to the US government and those in politics, such as future presidents and political figures, who may be influential during infectious disease outbreaks.

K.2 Julie Bishop

A number of Twitter users referred to news articles which mentioned the foreign Health Minister of Australia, Julie Bishop:

‘Ebola outbreak: Australia will not send health workers to West Africa – Foreign minister Julie Bishop says’ (K2.1)

*‘Australia won’t be sending health workers to West Africa – Julie Bishop is a Heartless B*tch’ (K2.2)*

The articles related to Julie Bishop were mostly neutral; however, a user suggested that Julie Bishop lacked a heart for not sending health workers to West Africa. There may have been a number of legitimate reasons for Australia to make the decision not to send health workers to West Africa. However, these reasons may not have been available to the Twitter users that were tweeting at the time.

Users on Twitter may make impulsive judgements and comments without having all available information.

K.3 Criticism of governments

This theme details critique towards both governments, in general, and the US government, in particular, rather than towards individuals. For instance, users were critical of how little money the US government had set aside to fight Ebola. Some users criticised the US government for fear-mongering. Other Twitter users criticised Australia's decision to not send health workers to fight Ebola, as indicated above, and there were some users who criticised the Liberian government:

'U.S Government gives over half billion to fight ISIL, however government only gives $8 million to fight Ebola – how stupid' (K3.1)

'Our stupid American government is fearmongering for war profiteers and fascists, yet a real threat to us all – Ebola spreads.' (K3.2)

'They say they have no money or manpower to fight Ebola – yet they have money for the war in Iraq!' (K3.3)

The majority of tweets in this sub-theme directed critique towards the US government. Specifically, users were criticising the lack of funds that had been allocated to tackle Ebola, as well as noting that the government had been fear-mongering and distracted from fighting Ebola.

A number of critical tweets were directed at the Australian government:

'Look at how foolish Abbott and co have been by not sending support to help Ebola but have send bombs and guns to fight Ebola' (K3.4)

'Strange decision – if the Ebola virus spreads then isn't the government going to look dumb?' (K3.5)

It appears that there was a tendency to criticise governments if they were perceived to provide inadequate support to West Africa. One Twitter user criticised the Liberian government:

'So many people are angry and feel betrayed in Liberia. People feel that their government has let them down' (K3.6)

Governments may be held accountable during infectious disease outbreaks and, if they do not deploy enough relief efforts, these criticisms may increase. In the case of the US government, it appeared that users would primarily criticise the lack of funds that had been allocated. In the case of Australia, the criticisms surrounded the decision by the government to not send troops to West Africa.

L. Humour and Sarcasm

Some tweets were sarcastic in nature and/or contained humour. These included sub-themes of sarcasm, humour, zombies, zombie apocalypse, and conspiracy theories. It is important to note that there was some cross-over to other themes and sub-themes for the *I.1 Sarcasm and Irony*, and *I.2 Humour* categories.

Table 5-11 Sub-themes of significant news stories

Theme	Sub-themes
L. Humour and Sarcasm	L.1 Sarcasm and Irony
	L.2 Humour
	L.3 Zombies
	L.4 Zombie Apocalypse

L.1 Sarcasm and Irony

The definitions of sarcasm and irony used in this study are outlined in H.5 sarcasm and irony. However, tweets did not have to fulfil the definitions completely and could have expressed a view in a sarcastic tone or been partially ironic. Illustrations of tweets coded within this category are provided below:

'Whoever favourites this has Ebola' (L1.1)

'Mother asked me what Ebola was – I replied it is a new track of my mixtape' (L1.2)

'Shame our concert is like Ebola' (L1.3)

'I bet you could catch Ebola from sex with a monkey' (L1.4)

'My father is the only person to have contracted Ebola from semen' (L1.5)

'My squad has the Ebola' (L1.6)

'Wonder where I could score me some Ebola?' (L1.7)

Twitter users would make false claims for the purpose of irony, and others were sharing tweets that expressed views in a sarcastic manner. It could be argued that Twitter users were not taking the Ebola outbreak seriously, hence the high number of ironic and sarcastic tweets posted. Users may also have been playing up to their followers by sharing posts for humorous purposes. Social media platforms are known to attract ironic and sarcastic comments, and as such this may be a trend of social media rather than a feature of the Ebola virus (Zappavigna, 2012).

<u>L.2 Humour</u>

Illustrations of tweets in the humour category are provided below:

'What did the fox say? Your Grannies got Ebola!' (L2.1)

'What meal would you like to cook with your ex...Ebola' (L2.2)

'Haha Nudes or Ebola…Laughing so hard' (L2.3)

'More beefs around this part than the Ebola' (L2.4)

'When you tweet about Ebola yet she keeps sucking' (L2.5)

As was noted in the sarcasm and irony sub-theme, Twitter users were sharing tweets that expressed humour and may not have taken the Ebola outbreak situation seriously. Moreover, Twitter users could have been over-exposed to the coverage of the Ebola outbreak. Therefore, users may have wanted to defuse the outbreak situation. There is a field of research known as 'Internet humour', which looks at humour as it occurs in social media. Linguistic researchers have argued that humour on social media is a form of showing solidarity with one-another (Zappavigna, 2012). Therefore, one explanation for the large number of tweets expressing sarcasm and humour could be that this was a way for users to come together to display solidarity.

<u>L.3 Zombies</u>

A number of Twitter users referred to zombies in the context of the Ebola outbreak. For instance, users would make a link between the Ebola outbreak and zombies in a humorous context, some users would state that there were zombies in West Africa, that Ebola was mutating to turn humans into zombies, and that people should purchase weapons for when zombies would rise after having died of Ebola. Illustrations of tweets coded in this theme are provided below:

'Night everyone – remember to not let Ebola turn you into a zombie' (L3.1)

'Did anyone else hear that there might be zombies in West Africa?' (L3.2)

'Wish this trend of Ebola Zombies would end' (L3.3)

'My bro said to me that Ebola zombies being somewhere' (L3.4)

'OMG, Ebola is causing zombies to happen' (L3.5)

As mentioned in section 2.6.2, zombies are often depicted in narratives found in Hollywood movies, television programmes, and video games. These narratives may begin with the premise of an infectious disease that is out of control and which is spreading across the world. These narratives may be the only reference point for infectious diseases for Twitter users who were tweeting during the outbreak. Therefore, Twitter users may reference zombies for this reason. Twitter users may also have been referencing zombies as a form of satire.

Other tweets referred to weaponry as a solution to zombies, as illustrated below:

'Bro let's find this mystery box! We need to take care of these Ebola zombies fast' (L3.6)

'Hearing about Ebola turning people into zombies, it is time to buy guns!' (L3.7)

In tweet *L3.6,* the Twitter user refers to a mystery box, which alludes to a box of weapons in the Call of Duty video game series. The theory behind the tweets is that when eventually the Ebola outbreak was to mutate into a virus which can turn humans into zombies, then users would use weaponry to eliminate them. This is a reference to popular culture because, in zombie films, ordinary people would arm themselves with weapons in order to eliminate zombies. Many video games also have this as premise, for example, the Call of Duty Zombies video series.

L.4 Zombie Apocalypse

Some Twitter users referred to the zombie apocalypse in the context of the Ebola outbreak. For example, a number users referenced a news article on the possibility of a zombie apocalypse which had been was published by the UK newspaper *The Mirror*. Illustrations of tweets in this category are:

'OMG, it looks like the zombie Apocalypse is here [URL link to zombie article]' (L4.1)

'We are on RED ALERT, US government preparing for the zombie apocalypse' (L4.2)

'Ebola victims are rising from the dead? We must prepare for the Zombie Apocalypse' (L4.3)

'Say hello to the Zombie apocalypse – aren't we so fu**ed' *(L4.4)*

'If there was ever going to be a zombie apocalypse – it will be from Ebola' *(L4.5)*

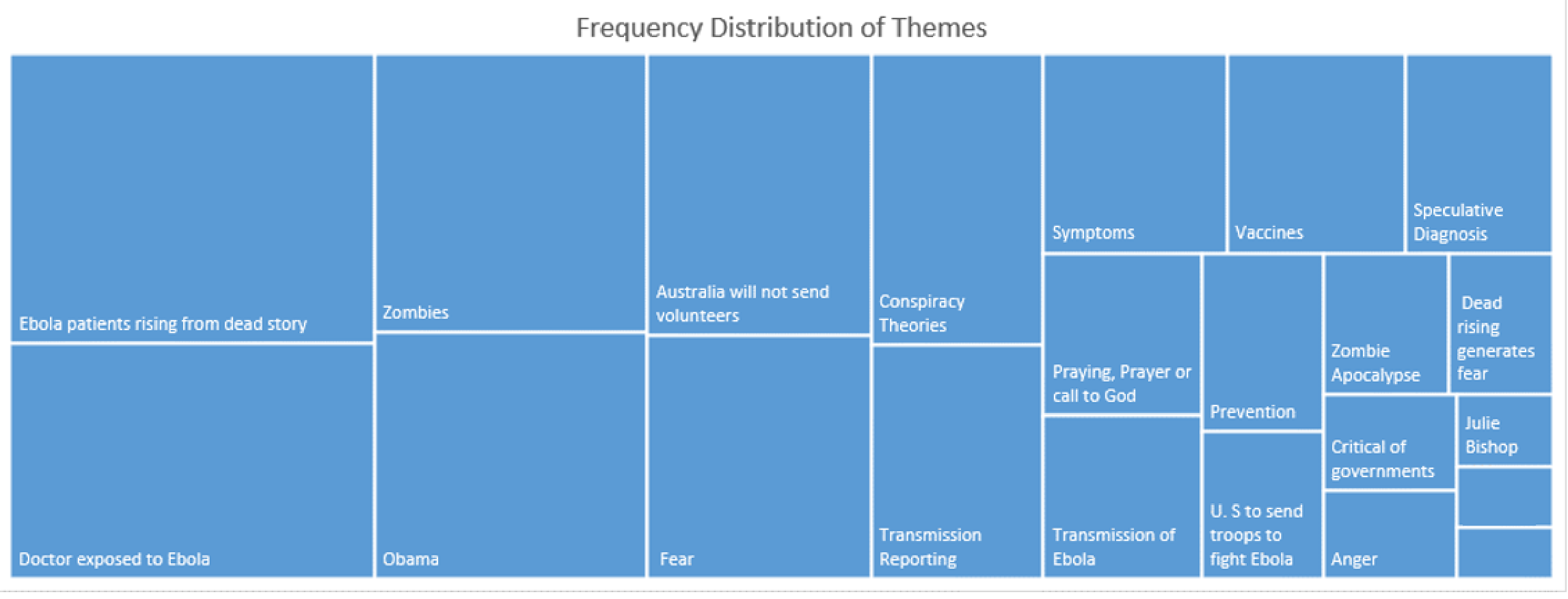
'Who ever thought that a zombie apocalypse was just something in video games and television? That shit is real' (L4.6)

Tweet *L4.5* may have been inspired by the symptoms of Ebola which are very gruesome and similar to that of zombies that are shown in television. Twitter users between the age of 18 and 24 (Pew Research Center, 2016) may be exposed to zombie narratives from a young age, for example, in the video games they play and the movies and television programs they watch. Therefore, Twitter users may bring up the possibility of a zombie apocalypse because it is something that they already have on their minds. Hence, when they hear about an infectious disease outbreak, their imagination may drift to these narratives. This can shape their understanding of the outbreak. However, it is difficult to know (without asking directly) whether the users genuinely believed that there was going to be a zombie apocalypse. To the best of the author's knowledge, this is the first evidence-based research to uncover these results (i.e. that Twitter users were referring to zombies and a Zombie Apocalypse in the context of the Ebola outbreak). The influence of popular culture on understanding infectious disease outbreaks is further explored in section 7.8.

5.7 Frequency Distribution of Themes

Figure 5-3 below is a tree map which displays the themes identified in this study. The theme of general discussions, sarcasm and humour were removed in order to generate the tree map below as due to their size they would over-shadow the other sub-themes. The size of the boxes represents the frequency of themes.

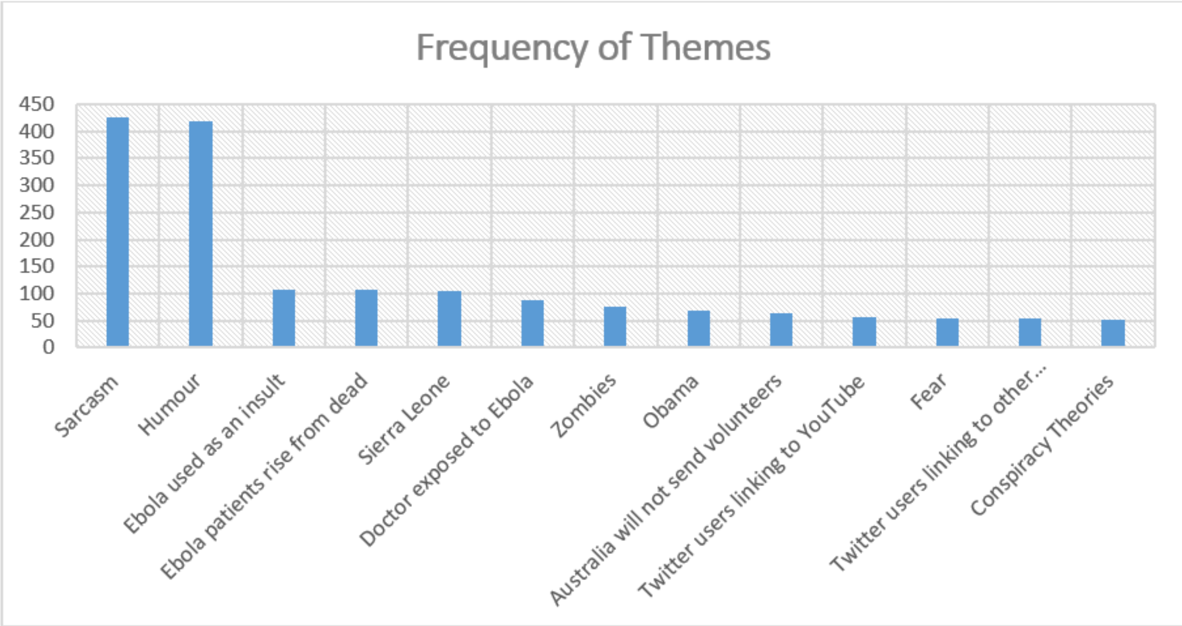
Figure 5-3 Tree map displaying the frequency distribution of sub-themes



*Note: 1 = News Story uses terrorism analogy 2 = Fear of Travel

It is interesting to note that a number of news stories dominated discussion on Twitter during this time period, for example, the story that Ebola patients had risen from the dead, that a US doctor had been exposed to Ebola, and the news that Australia would not send volunteers to West Africa. Figure 5-4 below is a bar chart which displays the frequency distribution of themes with more than 50 tweets assigned to them.

Figure 5-4 Frequency distribution of themes



It can be seen that the two most frequently occurring sub-themes were sarcasm (425/7%), and humour (418/7.35%). The third most frequently occurring sub-theme was users who used Ebola as an insult (108/1.95%), and the fourth most frequently occurring sub-theme was related to the news story that Ebola patients had risen from the dead (107/1.9%).

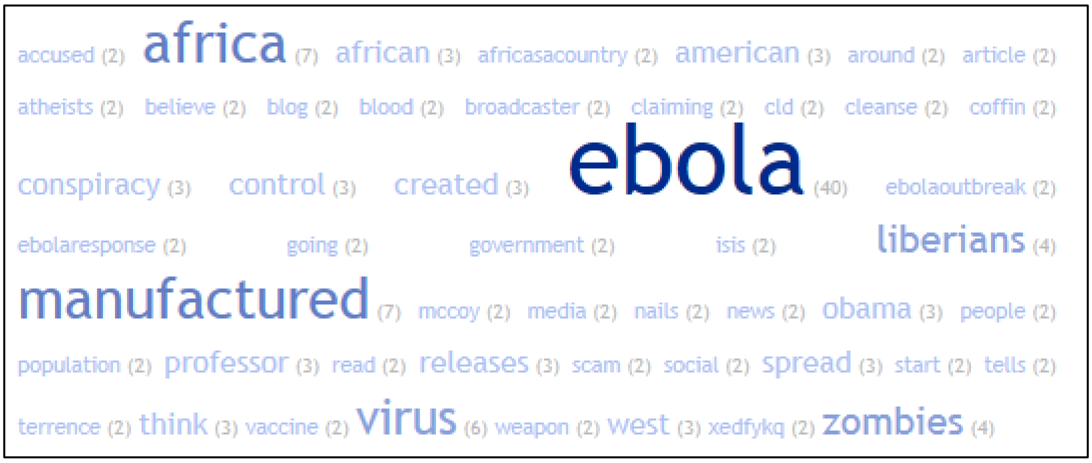
5.8 TAG Clouds of Sub-Themes

This section looks at the most frequently occurring words among the sub-themes. The application used in generate the word cloud was Tag Crowd (Steinbock, n.d.). The stop words (i.e. words that were not included within the word clouds) consisted of 'co', 'http' and 'rt'. Within the TAG clouds the words that appear larger are those which occurred most frequently within the sub-theme. The TAG clouds also contains the frequency of occurrence associated with each of the words. The figures below were created by extracting nodes, adding stop words (i.e. words such as Ebola), and then visualising the most frequently occurring words in Tag Crowd.

Figure 5-5 TAG cloud of zombies

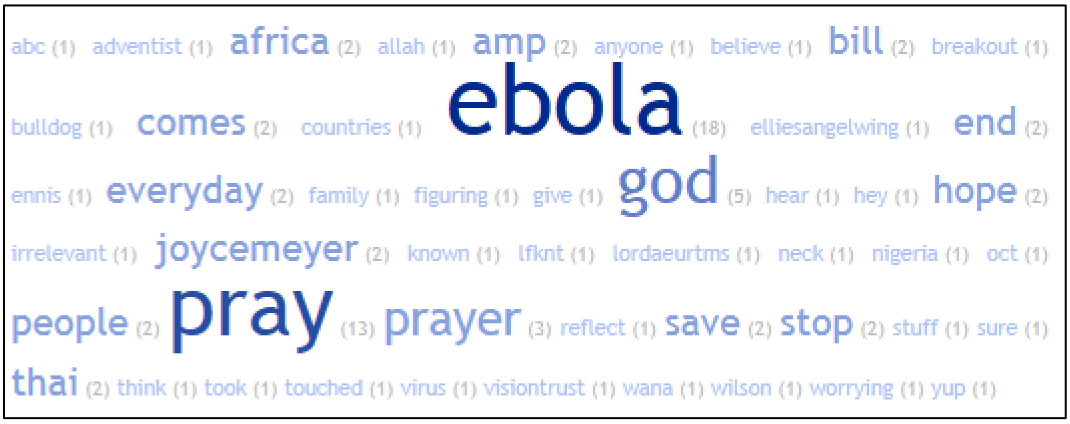
It is interesting to see words such as 'zombies', 'dead' and 'resurrected' appear among the tag cloud as well as the word 'real'. Twitter users were referring to how the dead were rising, and that zombies were real, because people were coming back to life.

Figure 5-6 TAG cloud of conspiracy



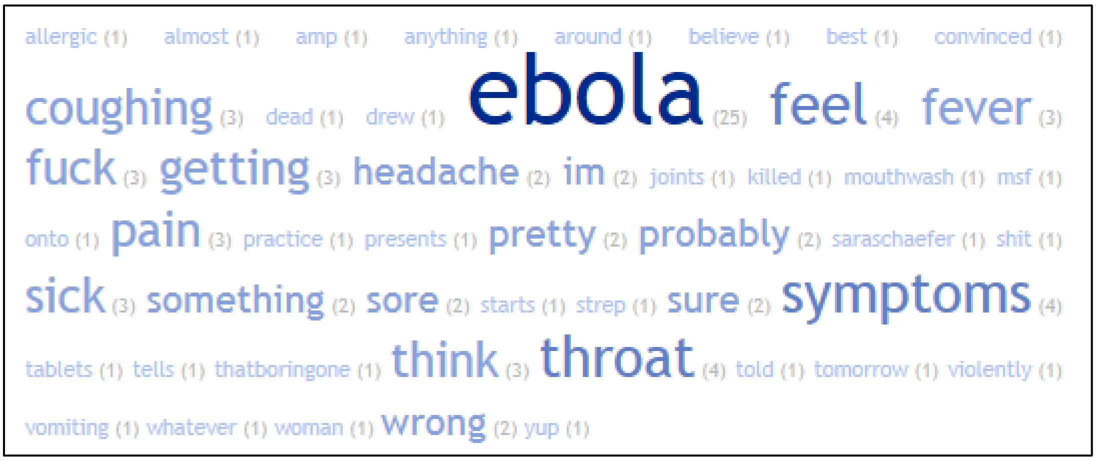
It is interesting to observe the word 'manufactured' appear among the frequently occurring words, although this only arose on seven occasions. Various conspiracy theories were shared during the outbreak. It appears that the theory that Ebola was manufactured was among the most popular.

Figure 5-7 TAG cloud of praying



Words such as 'Ebola', 'pray', 'god' and 'prayer' were among the most frequently occurring. Twitter users called other users to pray, and claimed that prayer would help with the Ebola outbreak. It is interesting to observe 'joycemeyer', a Christian charity, appear within the word cloud.

Figure 5-8 TAG cloud of symptoms

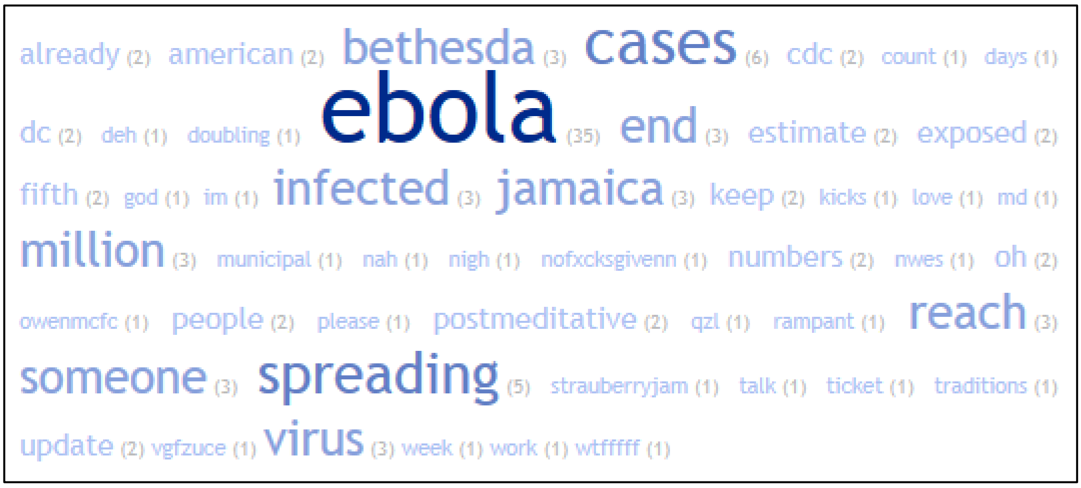


'Sore', 'throat', 'pain', 'coughing', 'headache' and 'fever' are among the most frequently occurring words. These words were written by users who were experiencing these symptoms and who feared they had been infected with Ebola.

Figure 5-9 TAG cloud fear of Ebola

Words indicating fear such as ‘scared’, ‘scary’, ‘gonna’, and ‘die’ appear within the most frequently occurring ones. There are also swear words such as ‘f**king’ which may have been used out of fear.

Figure 5-10 TAG cloud of transmission reporting of Ebola



It is interesting to observe geographical locations such as Bethesda and Jamaica appear among the most frequently occurring words within this sub-theme. These were locations where Ebola had been reported, or which were in the news. Words such as cases, infected, spreading, and teach also appear, since locations were mentioned when Twitter users commented on the spread/reach of Ebola.

5.9 Validity and Reliability of Results

5.9.1 Test-Retest Reliability

Test re-test reliability was outlined in Methodology Chapter 3, section 3.14.2. This test involves the same individual coding the data after a short period of time in order to compare results. Therefore, the author who coded the initial study (WA) re-coded a sub-set of data after a three-month period in order to assess test-retest reliability. The percentage agreement was 99.94%, and $\kappa = 0.87$, which is at a substantial level (McHugh, 2012).

5.9.2 Intercoder Reliability

Intercoder reliability was outlined in Methodology Chapter 3, section 3.10.5. A coder independent to the research project who was a doctoral candidate with a Master's degree was asked to code a subset of tweets. Intercoder reliability percentage agreement was 99.93% and $\kappa = 0.59$, which is at a moderate level (McHugh, 2012).

5.10 Summary of Twitter Dataset

For this case study, because tweets were sent during 2014, it has been possible to obtain further information on the sample of tweets. In total over this time period a total of 499,546 tweets were sent and received by 276,613 users, and 254,382 (51%) of tweets were original, 245,164 (49%) of tweets were retweets, and 285,450 tweets (57%) contained links. The majority of tweets were from the United States (54.9%), and the United Kingdom (4.8%) and a full summary of country locations for this dataset is summarised in Table 5-12. The location information of users that were tweeting was taken from the biography of the Twitter users. This is possible to obtain because, when a Twitter user registers for a Twitter account, they are asked information about their home country and city, and this is often displayed on a user's biography.

Table 5-12 Location of users tweeting

Country	Number	%
United States	127,328	54.9
United Kingdom	11,025	4.8
Nigeria	10,228	4.4
Canada	9,198	4.0
Brazil	5,706	2.5
France	4,570	2.0
South Africa	3,616	1.6
Spain	3,586	1.5
Indonesia	3,433	1.5
Australia	3,060	1.3

Table 5-13 Popular regions users were tweeting from

Region	Number	%
Texas	23,030	13.4
New York	12,968	7.5
California	12,745	7.4
England	7,782	4.5
Florida	6,258	3.6
Washington D.C.	5,400	3.1
Illinois	4,029	2.3
Lagos	4,014	2.3
Ontario	3,903	2.3
Georgia	3,802	2.2

As shown in the table above, the locations from where tweets were most frequently shared were Texas (13.4%), New York (7.5%), and California (7.4%). One reason people may have sent a large number of tweets from Texas is because a case of Ebola was diagnosed in this region. Table 5-14 below displays the cities with the most tweets.

Table 5-14 Ebola cities with most tweets

City	Number	%
Dallas	5,301	6.5
London	3,780	4.7
New York City	3,265	4.0
Los Angeles	3,168	3.9
Chicago	2,720	3.3
Houston	2,618	3.2
Atlanta	2,586	3.2
Austin	2,224	2.7
Toronto	2,171	2.7
Paris	1,792	2.2

The majority of tweets, at the city level, were derived from Dallas (6.5%), London (4.7%), and New York City (4%). Table 4-15 below displays the frequency of languages within this time period.

Table 5-15 Language of tweets

Language	Number	%	
English	400,506	80.2	
Spanish	25,157	5.0	
French	21,967	4.4	
Portuguese	11,586	2.3	
Italian	8,333	1.7	
German	7,300	1.5	
Indonesian	5,369	1.1	
Turkish	4,194	0.8	
Dutch	3,167	0.6	
Norwegian	1,242	0.2	

English was by far the most used language in tweets (80.2%), followed by Spanish (5%), and French (4.4%).

Table 5-16 Frequency of Ebola hashtag

Hashtag	Number	% of total dataset
#ebola	109,509	21.9
#breaking	13,609	2.7
#liberia	4,644	0.9
#dallas	4,310	0.9
#ebolaoutbreak	3,546	0.7
#news	3,157	0.6
#cdc	3,110	0.6
#sierraleone	2,719	0.5
#health	2,290	0.5
#tcot	2,166	0.4

Table 5-16 displays the most frequently occurring hashtags during the analysed interval of time (28th and 29th September 2014), whereby #ebola (21.9%) was the most popular hashtag followed by #breaking (2.7%), #liberia (0.9%), and #dallas (0.9%).

Table 5-17 Most retweeted tweets

No	Tweet	Country	City	Retweet Count
1	When you only survive during an Ebola outbreak and no one else [Link to Vine* of an African Child dancing]			733
2	Troops from the US arrive in Liberia ready to fight Ebola [Link to article by Wall Street Journal]	United States	Brooklyn	287
3	Omg! Hell! Turns out zombies are real! WTF [Link to Mirror article]			213
4	Here are how we are helping to fight Ebola [Link to UNICEF website]			204
5	Here is everything you will need to know about Ebola [Link to YouTube]	United States	San Bruno	197
6	Just don't get infected by Ebola or feelings for that matter, and you will be OK.			161
7	Help to stop Ebola must come within weeks or the scale of resources needed will be unimaginable. [Link to YouTube video]	Sweden		138
8	The Red Cross dead-body team are some of the most feared in Liberia [Link to Time news article]			134
9	Met someone from Nigeria who was fighting Polio, he is back now in to fight Ebola in Liberia	South Africa	Johannesburg	132
10	NIH will provide care for the US doctor who had obtained Ebola.			129

*Note = Vine is a video sharing platform

Table 5-17 above displays the most retweeted tweets from the dataset. The most popular tweets, tweet 3 and tweet 6, belong to the *humour and sarcasm theme.* Tweets 2, 4, 5, 7, 8, 9 and 10 belonged to the *general commentary* theme.

5.11 Identifying Influential Twitter users

Identifying Influential Twitter users was performed by extracting a sample of the network from 29/09/2014 (09:00 to 12:00). Table 4-18 below provides an overview of the network metrics for the data that was sampled. There were a total of 7078 nodes within the network, i.e. the number of Twitter users, and there were a total of 5796 edges, i.e. the connection between users.

Table 4-18 below provides a ranking of Twitter users by InDegree.

Table 4-18 Ranking Twitter users by InDegree

Node	InDegree	Account Description
A	743	Member of Public
B	248	Member of Public
C	246	Member of Public
D	119	Time magazine
E	112	Spam

Node A – Member of the public

The user was a member of the public who shared a vine video of an African child who appeared to be performing a dance in a village with the caption 'when you the only one who survives ebola in your village', and this was shared widely on Twitter.

Node B – Member of the public

A user who works for the wall street journal as a photographer shared picture of US troops when they arrived in Liberia to fight Ebola. The picture was taken by node C, and this user was tagged by node B in the photograph.

Node C- Member of the public

This user is a photographer and was tagged by node B in a tweet, and therefore received many inbound mentions and connections. However, this user did not tweet during the Ebola outbreak themselves.

Node D- Time magazine

Time magazine would share a news article which would indicate that the most feared people Liberia, an Ebola affected area, are the Red Cross dead body management team.

Node E- Spam

Node E appeared to be a spam and/or bot account that tweeted about the Liberian health minister quarantining themselves after their office assistant dies of Ebola. This account was tagging itself and other accounts that appeared to be spam accounts. The tweets from this user had no likes and/or engagements.

Table 4-19 below provides a ranking of Twitter users using the metric of OutDegree.

Table 4-19 Ranking Twitter users by OutDegree

Node	OutDegree	Account Description
A	19	Nurses for America
B	19	News account
C	19	Spam account
D	18	Spam account
E	17	Spam account

Node A – Nurses for America

This Twitter account was tweeting widely during the outbreak and was sharing their opinions on the events that were taking place and engaging in Twitter chats. For instance, the account was critical of the CDC and their quarantine efforts.

Node B – News Account

Node B was a news account which was frequently tagged by other Twitter users in Ebola related discussions, however, this Twitter account was not found to be tweeting about Ebola themselves.

Node C –Spam Account

This account appeared to be spam based and was replying to users with links to videos, and news articles in the hope that people would click on the links which is a tactic known as link-baiting.

Node D – Spam Account

This Twitter account appeared to be a spam account and was replying to Twitter users who were mentioning Ebola with links to articles, i.e. link baiting.

Node E – Spam Account

This Twitter account appeared to be a spam account that was tweeting during this time.

5.12 Evidence of themes across the outbreak

Section 4.15 provided a justification for searching for themes across the outbreak. In this study Twitter's advance search feature was used in to tweets from across the Ebola epidemic from when Google Trends showed that there was an increased interest in web search queries, and this was from April 2014 to August 2014. The themes were selected based on whether they were not specific to the 2-day time period, were not reported in previous literature, and which were searchable by using keywords. By running a number of searchers on Twitter's advance search tools, tweets were found from across the epidemic (April to August) that expressed opinions which were identified in the following themes: E.2 Fear, H.11 Conspiracy Theories, H.10 Downplaying Ebola risk, J.3 Zombies, H.10 Downplaying Ebola risk, and H.5 Western Privilege.

Table 5-20 Table of themes from across the outbreak

Themes	April	May	June	July	August
E.2 Fear	x	x	x	x	x
H.11 Conspiracy Theories	x	x	x	x	x
E.1 Obama	x	x	x	x	x
J.3 Zombies	x	x	x	x	x
H.10 Downplaying Ebola risk	x	x	x	x	x
H.5 Western Privilege	x	x	x	x	x

This provides evidence for Twitter users expressing views associated with some of the themes across the outbreak of Ebola from May 20009 to November 2009. That is to say, Twitter users were sharing tweets related to fear, were discussing conspiracy theories, referring to Obama, referring to zombies, downplaying the risk of Ebola, and mentioning the concept around Western Privilege throughout the outbreak period.

5.13 Discussion

5.13.1 Key Findings

The main finding was that discussions on Twitter involving Ebola revolved around eight key themes (percent calculated from total number of tweets coded):

- Emotion and feeling (113/2.60%)
- Health information (192/4.5%)
- Significant news stories (282/6.60%)
- General commentary (2311/54.0%)
- References to official organisations (75/1.80%)
- References to West African cities and or regions (181/4.2%)
- Political references (88/2.05%)
- Humour and sarcasm (1046/24.41%)

In Discussion Chapter 7, section 7.3.10, the results of this study are compared to previous research, and 41 or more potential themes emerged that, to the best of the author's knowledge, have not been reported in previous empirical studies. This demonstrates the potential of in-depth methods to uncover more insight into content shared on Twitter during infectious disease outbreaks.

One key finding was that the most frequently occurring theme was that of sarcasm and humour, as Twitter users on some occasions would defuse the outbreak situation by sharing a number of humorous tweets and posts. Sub-themes comprised of humorous, ironic and or sarcastic tweets. These sub-themes were as follows: sarcasm (425/7%), humour (418/7%), zombies (77/1.3%), zombie apocalypse (18/0.3%), and Ebola used as an insult (108/1.9%). The symptoms of Ebola appear to be very unpleasant and for some of the users tweeting popular culture may have been the only reference point leading to users to mention narratives observed in popular culture deceptions of zombies.

It was also found that a news story suggesting that Ebola victims had risen from the dead was shared by Twitter users frequently (107/1.9%) during the outbreak. The story also had the potential to cause fear among Twitter users and a specific sub-theme was created for tweets where users reported that the story had made them feel afraid. This sub-theme was labelled as '*Dead rising generates fear'* (15/0.26%). This news story had not factual basis and it is also unclear to what extent users on Twitter genuinely believed that Ebola had caused the dead to rise. However, the story is dangerous because in parts of West Africa there is a distrust of western medical science and a focus on non-traditional healing (Ѐreeman and Motsei, 1992).

In addition, there were a number of further news stories appeared to be significant during this time period (282/5%) and were shared frequently on the platform. These were as followed: Australia will not send volunteers (64/1.1%), US to send troops to fight Ebola (18/0.31%), news story uses terrorism analogy (6/0.10%), Doctor exposed to Ebola (87/1.53%), and FDA warning over fake drugs (30/0.52%).

Another interesting finding was that Twitter users would link to YouTube (56/0.9%), and also to other tweets (55/1%) when tweeting about the Ebola outbreak. It was also interesting to find that Ebola had the potential to cause fear among Twitter as a number of sub-themes was identified in this respect, for example fear (55/0.9%), fear of travel (5/0.1%), dead rising story generates fear (15/0.26%). Conspiracy theories, as well as news stories reporting on conspiracy theories, were shared by Twitter users (51/0.9%). These theories suggested that the US had spread Ebola as a form of population control, and that Ebola was a failed bio-weapon. These findings have a number of implications for public health informatics, and are considered in the next section.

When examining influential Twitter users by InDegree i..e, users being mentioned the most, it appeared that those users who were most influential would share humorous content. When examining users by OutDegree, it appeared that users who would tweet the most their accounts were centred on spam.

5.13.2 Implications of Results for Practice and Policy

The results relating to users expressing fear and anger in their tweets may be of interest to health authorities responsible for disseminating information during infectious disease outbreaks. Health authorities are known to disseminate information aimed at reducing stress among the general public during the height of an outbreak. Therefore, in potential future cases of Ebola, health authorities could consider sharing stress reduction and coping strategies when dealing with the threat of an infectious disease.

Health related information such as symptoms, transmission, vaccines, and prevention information may also be of interest to health authorities disseminating information during infectious disease outbreaks. Information could potentially be used to feed into the content that would be shared by health authorities via their Twitter accounts. For example, a tweet or a series of tweets could be sent out outlining how Ebola is transmitted, its key symptoms, and information related to vaccines. A benefit of using Twitter data for this purpose is that the data are available in real time and, as such, would allow health authorities to rapidly assess the information needs of users around transmission. Moreover, during the outbreak of a deadly infectious disease, real-time, up-to-date information from the general public could be vital in the battle against further spreading.

The results surrounding conspiracy theories could be of potential interest to the US government as well as health authorities worldwide. False information that goes uncorrected can have very serious consequences. For example, during the Ebola epidemic of 2014 there were riots in Guinea due to rumours that health workers were transmitting virus to locals (Jansen, 2014). In light of this, health authorities may wish to monitor discussions on Twitter, and rapidly correct potential misinformation. Moreover, it must be noted that, within this sample, no tweets from health authorities were sent to counter conspiracy theories and false information. Therefore, in future cases of infectious disease outbreaks, health authorities should consider releasing statements and information regarding the validity of the conspiracy theories.

Some users were very concerned when Ebola was reported in close proximity to them, and some people tracked the location of the outbreak. These findings could be of potential interest to health organisations as they could disseminate information to users in the areas affected using targeted advertising provided by social media platforms. Twitter and Facebook offer the ability to create adverts which only appear to users who reside in specific geographical locations. Therefore, in instances such as this, it would be possible for health organisations to target specific areas in order to provide advice and information to the general public.

Governments are expected to bring together resources and tackle infectious disease outbreaks such as Ebola. When it appears that governments are not doing enough they are exposed to criticism. However, misunderstandings can circulate among the general public about how a government is responding to an outbreak. Governments could use this information and respond to these suggestions on Twitter.

Health authorities and organisations may wish to monitor the narratives that Twitter users refer to in an attempt to understand infectious disease outbreaks. For example, it was found that a number of Twitter users would refer to popular culture such as zombies and the potential of a zombie apocalypse. These narratives may lead some users to believe that diseases such as Ebola could affect the brain and that they could lose control of their bodies. During times such as this, health authorities could consider appealing for calm on the Twitter platform, or using popular culture to communicate public health messages.

5.13.3 Health Belief Model

The theoretical framework of the Health Belief Model was applied in the results above and was useful in understanding the potential motivation of Twitter users. Ebola has a high perceived severity as it has a high fatality rate and Twitter users were more likely to alter their behaviour in order to avoid contracting Ebola. Moreover, Twitter users could potentially be more afraid of Ebola in comparison to other diseases such as the Influenza virus. Furthermore, when looking at tweets related to travel a possibility which was considered was that as the perceived severity and seriousness of Ebola was high, users were likely to avoid travel to West Africa. This is because people may have wanted to avoid a potential negative health outcome (Ebola) by changing their plans (behaviour change). There were also cases of people cancelling scheduled trips to West Africa in order to avoid visiting affected areas. It was also found that people were using information from Twitter to track Ebola and to monitor whether it was approaching the region that they were residing in. When Twitter users were alerted of cases of Ebola in close proximity, they became more fearful, as their perceived susceptibility of Ebola would increase. As noted previously when applying the HBM to results derived from data on Swine Flu, a limitation of the model is that it may appear to over-simplify human behaviour. The utility of the HBM is further summarised in Chapter 7 Section 7.5 which offers an outline of where the model was specifically utilised and its usefulness. The Health Belief Model will be further applied and discussed in Chapter 6.

5.13.4 Results in context of Public Health Informatics

The results of this study can be situated within the field of Public Health Informatics, which is a sub-field of Health Informatics. This study was able to gain insights into the public views and opinions based on tweets at the height of the Ebola outbreak. As highlighted in Chapter 2, previous studies on infectious disease outbreaks have used quantitative methods of analysing Twitter data. This study is potentially the largest qualitative analysis of tweets related to the Ebola epidemic of 2014. The methodology employed in this study in relation to the sampling of data, and coding of tweets using thematic analysis, may be of interest to public health researchers, who could potentially utilise data from Twitter during infectious disease outbreaks and use qualitative methods to gain insight into tweets in real-time. This would allow public health officials to rapidly disseminate information on the platform during an epidemic. This study has also uncovered a number of interesting results surrounding false information, conspiracy theories, as well as the influence that popular culture can have on Twitter users. These may be of potential interest to public health researches and are areas that could be further researched.

5.14 Limitations

This study only examined a two-day period selected at the peak of the Ebola outbreak of 2014, and therefore there are limitations to the conclusions that can be drawn. However, section 5.12 did provide some evidence that some of the discussions taking place by Twitter users would occur throughout the outbreak. This is because the main themes and topics that may have been discussed on Twitter may have been altered. If other time intervals were examined, the main themes and sub-theme could emerge under different headings and sub-headings. It must also be noted that people posting on Twitter may behave differently online than they would do in real life and, therefore, some of the views that are expressed may be exaggerated. Moreover, Twitter data are not representative of the national offline population, so the results cannot be generalised. Additionally, by examining English-language tweets and excluding tweets in other languages the study is limited because it may not have captured the voices of those who were directly affected and/or in the front line in West Africa where Ebola was initially reported because they may have been tweeting in other languages such as French.

5.15 Summary

This case study sought to better understand the types of information that was shared across a two-day period during the peak of the 2014 Ebola outbreak. The study has resulted in a number of interesting and surprising findings, for example, in relation to the rate at which the discussion on Ebola revolved around a number of news stories. There were also interesting results regarding the rate at which Twitter users shared their emotions. Users discussed and shared content related to politics, health-related information, referred to popular culture and expressed humour and sarcasm in their tweets. Similar to Chapter 4 on swine flu, the results of this study, as well as the method of capturing and filtering data, may be of interest to public health researchers and organisations. The next chapter introduces a smaller empirical study which was undertaken in order to understand the information that was shared related to the Zika outbreak.

Chapter 6 Zika Study

6.1 Introduction

Whilst the original study was underway, an outbreak of the Zika virus occurred. Current Twitter data is available at no cost via the Gardenhose API. As such, an additional, smaller empirical case study focussing on the 2016 Zika epidemic was conducted and has been included, and the results of this are compared with those from Swine 'Flu' and Ebola in Chapter 7. It is important to note that, as the data were obtained using a different application, some aspects of the methodology and analysis necessarily differ from the other two studies. Section 6.2 provides background information to this study, section 6.3 provides an overview of key events during the Zika outbreak, section 6.4 provides a summary of the data collection procedures, section 6.5 provides the results of the study, section 6.6 provides the frequency distribution of themes, section 6.7 outlines a number of reliability measures that were taken, section 6.8 provides evidence for the themes of the study to have occurred through the outbreak period, section 6.9 discusses the results of this study, section 6.10 describes the limitations of this study, and section 6.11 summarises the results.

6.2 Background

The Zika virus (ZIKV) is a developing arthropod-borne virus (arbovirus) which was first isolated from a monkey from the Zika forest in Uganda in the late 1940s (Musso, Nilles, and Cao-Lormeau, 2014; Fauci and Morens, 2016). Zika sufferers tend to exhibit mild symptoms and may experience a slight fever or rash, and joint, muscle, or eye pain. Zika is transmitted from the bite of specific mosquito species carrying the infection. Zika can also be transmitted through sexual intercourse, and perinatal transmission has also been reported (Musso, Nilles, and Cao-Lormeau, 2014). There is no cure or anti-viral treatment for Zika virus infection, neither is there a vaccine. The Zika virus can spread from the womb to the child in pregnant women and leads to birth defects such as microcephaly (Brasil *et al.*, 2016).

6.3 Key Events During the Outbreak

Table 6-1 below is recreated from Kindhauser, Allen, Frank, Santhana, and Dye (2016) and provides an overview of key events taking place around the Zika epidemic from 30th January to the 1st of February which fall on and/or around the dates for when the tweets were retrieved.

Table 6-1 Overview of timeline

Date	**30 January 2016**	**1st February 2016**	**1st February 2016**	**2nd February**
Event	• Jamaica reports first case of locally transmitted Zika.	• The WHO declares the current Zika epidemic a Public Health Emergency of International Concern because of links to microcephaly and other serious neurological disorders.	• Cabo Verde (an island in the central Atlantic Ocean) reports there had been 7081 suspected cases from Sept 2015 to Jan 2016.	• Chile reports three cases of Zika for those returning from Brazil, Colombia, and the Republic of Venezuela.

6.4 Summary of Data Collection

The strategy for extracting the sample of tweets, as outlined in Chapter 3 Methodology section 3.14, is summarised below:

- The original dataset that was retrieved consisted of 749,131 tweets that were retrieved from January 31st to February 1st 2016 using Twitter's Gardenhose API. This time interval was selected because Google Trends Data showed an increase in Web Search queries at this time (Google, 2017).
- The dataset was retrieved using the Boston University Twitter Analysis Toolkit (Groshek, 2014), which has access to the Gardenhose API.
- The keywords used to retrieve data were 'Zika' (section 3.5.1 provides an overview of the challenges of retrieving data via keywords).

- Exact duplicate tweets were removed from the dataset at a 60% threshold, which led to 76,943 single tweets.
- Near duplicate tweets were then removed at a 60% threshold which led to 20,421 tweets
- Finally, a 10% simple random sample of tweets was taken (2042 tweets), and entered into NVivo for coding using the method of thematic analysis. A simple random sample ensures that there is an equal chance of selecting each tweet.

It is important to note that, as the data was obtained using a different application, the data was not enriched in the same manner as Firehose data on swine flu and Ebola. The next section provides an overview of the themes and sub-themes that emerged from the in-depth thematic analysis conducted on the dataset. Section 4.6 and section 5.5 contain a list of key changes to the Twitter platform from 2009 to 2016.

6.5 Results – Qualitative Analyses

Table 6-2 provides an overview of themes and sub-themes that emerged from the analysis of tweets. The percentage is calculated from the total number of coded tweets (n=1960).

Table 6-2 Overview of themes and sub-themes

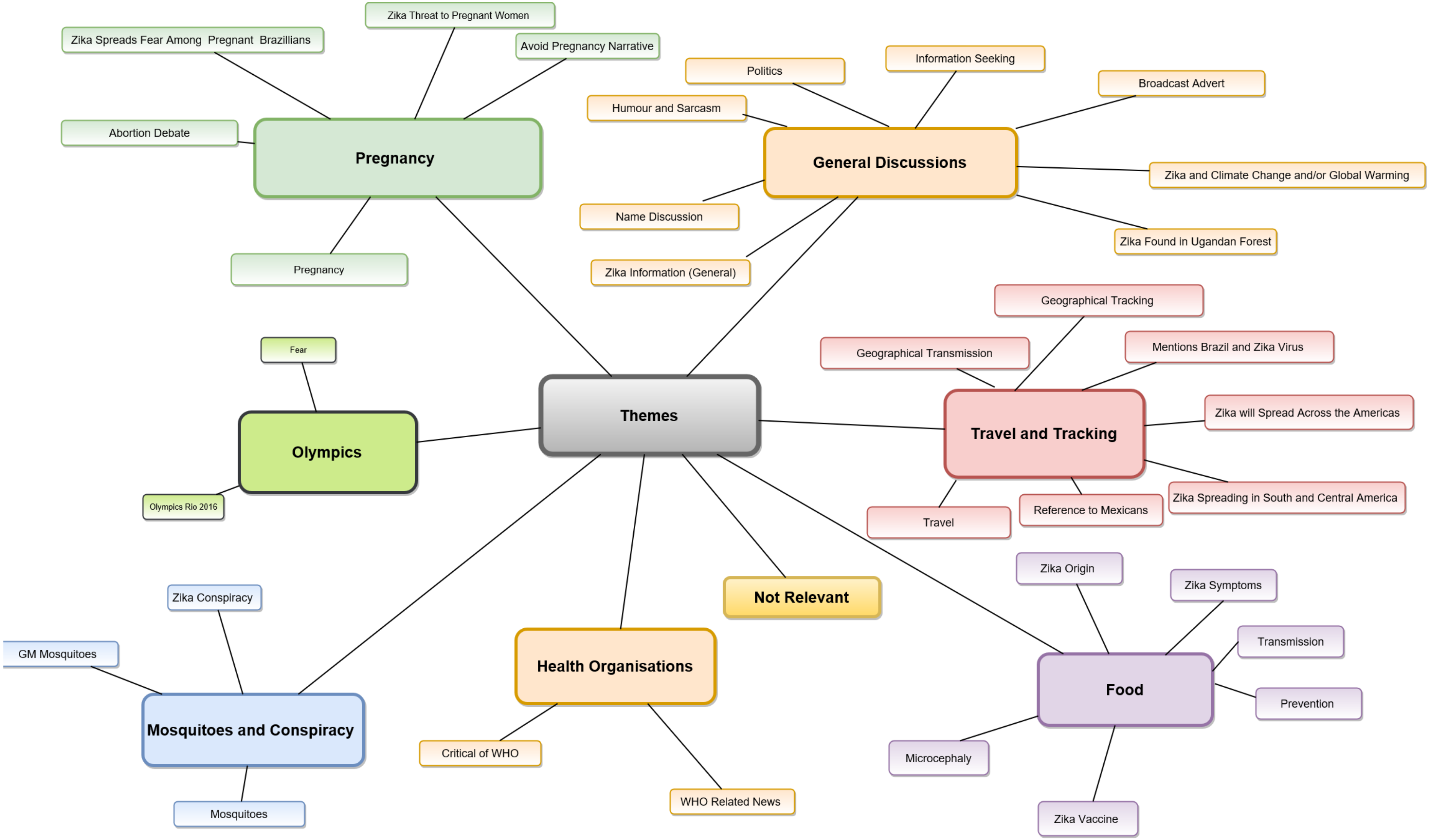
Theme (N/%)	Sub-themes (N)
1. Pregnancy (164/8.36%)	1.1 Avoid Pregnancy Narrative (12/0.61%) 1.2 Zika Threat to Pregnant Women (19/0.96%) 1.3 Zika Virus Spreads Fear Among Pregnant Brazilians (22/1.12%) 1.4 Zika Threat to Pregnant Columbians (29/1.47%) 1.5 Abortion Debate (47/2.39%) 1.6 Pregnancy (35/1.78%)
2. Olympics (75/3.82%)	2.1 Fear (19/0.96%) 2.2 Olympics Rio 2016 (56/2.85%)
3.Mosquitoes and Conspiracy (259/13.21%)	3.1 Zika Conspiracy (61/3.11%) 3.2 GM Mosquitoes (65/3,31%) 3.3 Mosquitoes (133/6.78%)
4. Health Organisations (365/18.62%)	4.1 Critical of WHO (7/0.35%) 4.2 WHO Related News (358/18.26%)
5. Health Information (160/8.16%)	5.1 Zika Origin (8/0.40%) 5.2 Zika Symptoms (10/0.51%) 5.3 Transmission (19/0.96%) 5.4 Prevention (34/1.73%) 5.6 Zika Vaccine (37/1.88%) 5.7 Microcephaly (52/2.65%)
6. Travel and Tracking (321/16.37%)	6.1 Zika Travel Advice (29/1.47%) 6.2 Zika Spreading Explosively in South and Central America (14/0.71%) 6.3 Zika will Spread Across the Americas (17/0.86%) 6.4 Mentions Brazil and Zika Virus (19/0.96%) 6.5 Geographical Tracking (21/1.07%) 6.6 Geographical Transmission (116/5.91%) 6.7 Travel (45/2.29%)
7. General Discussions (646/32.95%%)	7.1 Zika found in Ugandan Forest (15/0.76%) 7.2 Zika and Climate Change or Global Warming (18/0.91%) 7.3 Broadcast Advert (21/1.07%) 7.4 Information Seeking (26/1.32%) 7.5 Politics (26/1.32%) 7.6 Humour and Sarcasm (77/3.92%) 7.7 Name Discussion (124/6.32%) 7.8 Zika Information (General) (363/18.52%)

Seven prominent themes emerged from the dataset. These are listed below and are described in Table 6-2 and Figure 6-1:

- Theme 1: **Pregnancy**
- Theme 2: **Olympics**
- Theme 3: **Mosquitoes and Conspiracy**
- Theme 4: **Health Organisations**
- Theme 5: **Health Information**
- Theme 6: **Travel and Tracking**
- Theme 7: **General Discussions**

82 tweets were irrelevant to the Zika outbreak, and were removed during the data analysis phase.

Figure 6-1 Diagrammatical overview of themes and sub-themes (Zika)



1. Pregnancy

This theme contains tweets which contained references to pregnancy and which are based on the following sub-themes: a narrative suggesting women should avoid pregnancy, that Zika was a threat to pregnant women, the virus spreading fear among pregnant Brazilians, Zika was posing a threat to pregnant Columbians, and discussions around abortion.

Table 6-3 Sub-themes of pregnancy

Theme	Sub-themes
1. Pregnancy	1.1 Avoid Pregnancy Narrative
	1.2 Zika Threat to Pregnant Women
	1.3 Zika virus Spreads Fear Among Pregnant Brazilians
	1.4 Pregnant Columbians hit with Zika
	1.5 Abortion Debate
	1.6 Pregnancy

1.1 Avoid pregnancy narrative

A number of tweets contained links to articles from the mainstream media that noted women in Zika affected areas should avoid becoming pregnant:

'Doctors suggest to not get pregnant until further information on Zika is released [URL]' (1.1.1)

'Women who live in El Salvador have been informed to delay pregnancy till 2018 because of Zika [URL]' (1.1.2)

Zika is known to cause complications during child birth (Musso, Nilles, and Cao-Lormeau, 2014), the reason health authorities were recommending this course of action. The articles would also provide more general guidance on how to minimise the risk posed by the Zika virus.

1.2 Zika virus threat to pregnant women

The previous theme contained tweets providing specific advice to women and couples on avoiding Zika. Other tweets linked to articles providing more general warnings as well as specifically for pregnant women:

'The Zika Virus is a real genuine threat to pregnant women [URL]' (1.2.1)

'Read this article on the threat of Zika for pregnant women [URL]' (1.2.2)

Zika was more of a threat to women who were pregnant, and, as such led to a number of news media outlets sharing articles on this theme, which were tweeted by Twitter users.

1.3 Zika Virus Spreads Fear among Pregnant Brazilians

Several tweets linked to an article that suggested the Zika virus had caused concern among pregnant Brazilians:

'Zika virus spreads Fear among pregnant Brazilians [URL]' (1.3.1)

'Info on Zika: It has been spreading fear across pregnant Brazilians [URL]' (1.3.2)

The tweets which link to news articles could have induced fear, as they convey a sense the virus was spreading rapidly and causing concern among women from Brazil who were pregnant. The tweets and articles in this theme specifically mentioned Brazilians because there was a surge of Zika cases in Brazil (CDC, 2016).

1.4 Pregnant Colombians hit by Zika

A number of Twitter users also shared news of Colombian women who were infected with Zika:

'Over 100 pregnant Colombian women have Zika [URL]' (1.4.1)

'Colombia is now reporting that there are 100 confirmed cases of Zika [URL]' (1.4.2)

The tweets in this theme could have caused concern among citizens because it appeared that the Zika virus was infecting pregnant women in Colombia at a high rate. The articles referred specifically to Colombians because there were a number Zika cases in Colombia (CDC, 2016).

1.5 Abortion Debate

General news articles shared on Twitter highlighted similar debates around pregnancy:

'The Zika Virus has sparked a debate around abortion and it is not the first disease to do so [URL]' (1.5.1)

The tweet above links to an article that specifies how the mainstream media were aware the Zika outbreak had initiated a debate around abortion, partly because, in certain areas where Zika had spread, laws exist which prohibit abortion (Aiken *et al*., 2016).

This specific tweet also alludes to previous diseases that have led to similar abortion debates, and it is possible to draw parallels with earlier infectious disease outbreaks. In the 1960s, there were laws that prohibited abortion in the United States and, during this time, there was an outbreak of Rubella, which has the potential to cause complications in foetuses (Niswander, 1965). The debate in the context of Zika centred on whether abortion should be decriminalised in areas such as El Salvador (a small country in Central America), where illegal abortions are common place (Roa, 2016). A number of Twitter users also referred to the practicalities of requesting women to avoid becoming pregnant:

'In Zika affected areas women are being told to not get pregnant, yet, they have no access to contraception' (1.5.2)

'Wonder if people realise that the narrative of not getting pregnant actually needs rights and resources' (1.5.3)

'Hopefully not, but if Zika ever gets to Nigeria I wonder if we are going to legalise abortion for affected women?' (1.5.4)

The series of tweets above highlight some of the debates that were taking place around abortion. In *tweet 1.5.2*, the user highlights the controversy of advice to avoid pregnancy when contraception is not readily accessible, as is often the case in the low-income countries affected by Zika (Williamson, Parkes, Wight, Petticrew, and Hart, 2009). Twitter users such as the author of *Tweet 1.5.3* suggested that avoiding pregnancy requires certain women's rights, specifically the right to abortion. *Tweet 1.5.4* questioned whether Nigeria, which has strict anti-abortion laws, would legalise abortion (Iyioha and Nwabueze, 2016).

1.6 Pregnancy

A number of more general tweets referred to the concept of pregnancy:

'All you need to know if you are pregnant or are thinking about pregnancy' (1.6.1)

'Here is information related to Zika and pregnancy'(1.6.1)

The tweets in this theme often linked to news articles that arose because of the threat of Zika to pregnant women.

2. Travel and Olympics

This theme consists of tweets related to the Olympics in 2016, travel concerns, and tweets where Twitter users expressed fear towards the Zika outbreak.

Table 6-4 Sub-themes of travel and Olympics

Theme	Sub-themes
2. Travel and Olympics	**2.1 Olympics Rio**
	2.2 Fear

2.1 Olympics in Rio

The Zika outbreak took place during the 2016 Olympics, which necessarily became a topic on Twitter:

'Athletes in Rio have been put in a house arrest because of worry surrounding Zika' (2.1.1)

'There's no risk that Brazil government will cancel the Olympics due to Zika' (2.1.2)

The tweets in this theme highlight how infectious disease outbreaks can have negative consequences on sporting events and, more generally, events with an international audience. Citizens who were travelling to Brazil for the Olympic Games may have been aware of the threat of Zika. Moreover, in the context of the Health Belief Model, this may have led certain groups, such as pregnant women, who had a higher perceived risk severity, to alter their plans and behaviour. Scholars at the time noted how international gatherings in 2016 were also a risk for the spread of Zika, such as the Rio Carnival, the Hajj and the Umrah pilgrimage, which receive visitors from Latin America (Elachola, Gozzer, Zhuo, Memish, 2016).

2.2 Fear

There were a number of Twitter users who expressed a sense of fear and concern during the Zika virus:

*'This Zika Sh*t is out of control! ' (2.3.1)*

'Zika is here – God help us' (2.3.2)

'Zika sounds like a work of horrid fiction brought into our reality' (2.3.3)

The Zika outbreak was seen to be spreading rapidly and, as such, was likely to cause Twitter users to express concern and fear. There may have been a sense that the virus was out of control, and this was particularly frightening for those who were more vulnerable to Zika.

3. Mosquitoes and Conspiracy

This theme contained tweets which would discuss genetically modified mosquitoes, mosquitoes in general terms, and conspiracy theories around Zika.

Table 6-5 Sub-themes of mosquitoes and conspiracy

Theme	Sub-themes
3 Mosquitoes and Conspiracy	**3.1 GM Mosquitoes**
	3.2 Mosquitoes
	3.3 Zika Conspiracy

3.1 GM Mosquitoes

A number of Twitter users referred to genetically modified mosquitoes within their tweets:

'The current outbreak of Zika could have originated from genetically modified mosquitoes put into the air [URL] ' (3.1.1)

'Genetically modified mosquitoes that have genes to self-destruct could help to fight against Zika'' (3.1.2)

It has bbeen widely reported that mosquitoes transmit the Zika virus (Musso, Nilles, and Cao-Lormeau, 2014), and some users believed genetically modified mosquitoes were specifically to blame for the outbreak. These opinions originate from the open air release of genetically modified mosquitoes in north-east Brazil, a measure undertaken to help fight mosquitoes harbouring the Zika virus (Panjwani and Wilson, 2016). When Zika remerged in 2016, it was found to be most prevalent in the region where the genetically modified mosquitoes were released, leading to the conspiracy theories shared in *tweet 3.1.1* (Panjwani and Wilson, 2016).

3.2 Mosquitoes

A number of Twitter users referred to mosquitoes more generally within their tweets:

'Imagine how bad central India would be affected if mosquitoes were to arrive' (3.2.1)

'W*e could eradicate mosquitoes once and for all' (3.2.2)*

'I am confident I do not have Zika because a mosquito hasn't bitten me for some months' (3.2.3)

Twitter users expressed interest around the geographical spread of mosquitoes that transmit Zika. Certain users would also deem themselves to be safe from developing Zika if they had not been bitten by a mosquito, a finding which may highlight a gap in public knowledge about the disease and how it might be transmitted, and as such be of interest to health authorities.

3.3 Zika Conspiracy

A number of Twitter users shared and discussed conspiracy theories focussed around the Zika outbreak:

'Zika – it's a hoax and the cover up related to it continues' (3.3.1)

'Bill Gates created the Zika virus [URL]' (3.3.2)

'I wonder who really started the Zika virus' (3.3.3)

'Why would you invent a virus and not have the antidote?' (3.3.4)

As the tweets above demonstrate, the Zika outbreak led Twitter users to share and discuss a number of varied conspiracy theories. Some theories purported that Zika had been intentionally spread (as illustrated in *tweet 3.3.3* and *tweet 3.3.4*). Other conspiracy theorists would suggest that certain prominent figures had intentionally manufactured the Zika virus (as illustrated in *tweet 3.3.2*). Tweet illustration 3.3.2 could also indicate humour because Bill Gates is known for the development of computer software and 'virus' is also a term utilised in the computing industry.

4 Health Organisations

This theme is based on tweets which were critical of the WHO or tweets that shared news related to the WHO.

Table 6-6 Sub-themes of health organisations

Theme	Sub-themes
4. Health Organisations	**4.1 Criticism of WHO**
	4.2 WHO Related News

4.1 Criticism of WHO

A number of Twitter users were critical towards the World Health Organisation:

'In 2014 we were hoping for WHO to do something re: Ebola. They focused on electronic cigarettes, and now with Zika do something WHO!' (4.1.1)

'Difficult to take WHO seriously after they were warning us that bacon was dangerous' (4.1.2)

'The new smoking on screen guidelines would mean some classics are given the 18 rating – WHO should focus on Zika rather than come up with stupid ideas' (4.1.3)

Twitter users appeared to have been displeased with the WHO for a number of reasons, and complained that the WHO had failed to take appropriate action in previous infectious disease outbreaks such as the Ebola epidemic.

4.2 WHO-related News

A number of Twitter users shared tweets containing links to news statements shared by the WHO:

'WHO applies pressure on international community to eradicate Zika [URL]' (4.2.1)

'WHO ponders over a global health emergency announcement on Zika [URL]' (4.2.2)

'WHO notes the Zika virus is a public health emergency' (4.2.3)

'The WHO announce Zika a global health threat' (4.2.4)

The number of tweets citing information originating from the WHO demonstrates that the WHO was seen as an authority figure for news and updates during the Zika. Moreover, as outlined in section 6.3, during this time the WHO had declared the Zika outbreak to be a global health threat, which led to a number of tweets linking to this news.

5 Health Information

Table 6-7 Sub-themes of health information

Theme	Sub-themes
5. Health information	**5.1 Zika Origin**
	5.2 Zika Symptoms
	5.3 Transmission
	5.4 Prevention
	5.5 Zika Vaccine
	5.6 Microcephaly

5.1 Zika Origin

A small number of users on Twitter discussed the origin of the Zika virus:

'The mosquito-originating Zika originated in 1947 in the Zika forest less than 20 miles from Kampala (Uganda's capital) [URL]' (5.1.1)

'Zika was discovered in our country in 2012 after a 15-year-old got it' (5.1.2)

Interest in the Zika virus was sparked among the public because the disease had suddenly re-emerged and was reported to be spreading fast across the Americas (Fauci and Morens, 2016). Reports on Zika in the mainstream media may also have incited speculation on its origin. In the context of the Health Belief Model, people feel more secure around a disease if they possess information on how it operates and believe they can predict its future movements.

5.2 Zika Symptoms

A number of other Twitter users discussed the potential symptoms of the Zika virus:

'Zika has the potential to deform new-borns if the mother becomes infected with the Zika virus during pregnancy' (5.2.1)

'Zika can cause brain damage to babies according to the WHO [URL] ' (5.2.2)

In the context of the Health Belief Model, tweets in this theme were likely be of more interest to Twitter users with a high susceptibility to Zika. This is because, unlike certain other infectious diseases, such as swine flu and/or Ebola, Zika is the largest threat to specific groups of people, e.g., pregnant women.

5.3 Transmission

A number of Twitter users discussed how the Zika virus was transmitted:

'Mass gatherings have the potential to begin outbreaks like Zika [URL]' (5.3.1)

'Zika is spread via mosquitos – not surprising*' (5.3.2)*

'This is how the Zika virus is spread [URL]' (5.3.3)

Within this theme, Twitter users discussed the ideal conditions for the spread of Zika and shared links to news articles on how Zika was transmitted. Twitter users may have acted on this information, for instance, citizens may have opted to avoid large gatherings if they felt that this would reduce their perceived threat of attaining Zika.

5.4 Prevention

A number of Twitter users were sharing tweets indicating how to avoid contracting Zika:

'Zika is spreading fear – mothers are buying mosquito-resistant baby clothing' (5.4.1)

'Due to Zika – hope you are packing lots of mosquito repellent' (5.4.2)

'Bug zappers and ultrasonic repellents are not effective mosquito prevention methods for Zika – they are scam solutions' (5.4.3)

A number of different prevention techniques were mentioned, such as clothing for children (*tweet 5.4.1*). Other Twitter users referred to mosquito repellent (*tweet 5.4.2*), and certain Twitter users would note that some prevention products such as ultrasonic repellents would not be effective in protecting users against Zika. This theme highlights how Twitter may have served as a platform for the dissemination of information on what prevention products were effective and less likely to be effective. These results may be of interest to public health organisations who may wish to monitor public perceptions of prevention methods via Twitter.

5.5 Zika Vaccine

A number of Twitter users were sharing information related to the Zika vaccine:

'The vaccine for Zika is many months and possibly even years away [URL] '(5.5.1)

'Everyone around the world should now be aware of Zika, and I really hope there is a vaccine soon' (5.5.2)

'Vaccine for Zika might be years away – experts have warned [URL]' (5.5.3)

Tweets often contained links to articles commenting on whether there was a vaccine available to cure and/or protect against the Zika virus. This is a natural question and/or information need Twitter users may have expressed. At the time, there was an increased interest in Zika from media outlets.

5.6 Microcephaly

Twitter users referred to the condition of microcephaly in relation to Zika:

'Zika-based microcephaly rather than the virus is a public health emergency' (5.6.1)

'A causal relationship between Zika and microcephaly is alleged but not yet a scientific link' (5.6.2)

'How would the link between Microcephaly and Zika be proved?' (5.6.3)

'*Zika and microcephaly: scientists are looking into whether mosquito bites can lead to brain defects*' *(5.6.4)*

As the tweets above demonstrate, during the height of the Zika epidemic, Twitter users discussed the link between microcephaly and Zika. Microcephaly is a medical condition associated with Zika which affects children during birth and which causes the head of a child to be a smaller size than normal (Mlakar, Korva, Tul, Popović, *et al*., 2016). It appeared that tweets within this theme tweets were of a very different tone to others, much more serious in approach, and there was no attempt to trivialise the issues.

6 Travel and Tracking

Table 6-8 Sub-themes of travel and tracking

Theme	Sub-themes
6. Travel and Tracking	**6.1 Zika Travel Advice**
	6.2 Zika Spreading explosively in South and Central America
	6.3 Zika will Spread Across the Americas
	6.4 Mentions Brazil and Zika Virus
	6.5 Geographical Tracking
	6.6 Geographical Transmission
	6.7 Travel

6.1 Zika Travel Advice

A number of tweets involved sharing travel-related advice:

'You can read about our travel guidance for Zika here [URL]' (6.1.1)

'Zika by numbers: travel advice disseminated across the globe [URL]' (6.1.2)

'The U.S. has added four additional countries to the travel watch list [URL]' (6.1.3)

Zika is likely to have caused concern among citizens with plans to travel to Zika-affected areas. This may have been a stressful time for people who were unsure whether it was safe to travel. This concern may have led to information needs on the safety of travel, and tweets linking to news articles were likely to address potential information needs of the public.

6.2 Zika spreading explosively in Central and South America

A number of users shared links to news articles stating that Zika had been spreading explosively in Central and South America:

'Zika virus spreads explosively in South and Central America' (6.2.1)

'Zika virus spreading explosively now in over 20 countries in Central America' (6.2.1)

This theme highlights how certain geographical areas were perceived to be a larger threat which, according to the Health Belief Model, could raise the perceived severity of risk to individuals in these areas. It is interesting to note the vocabulary used, i.e. 'explosively', which transmits a sense of alarm and implies the outbreak was out of control.

6.3 Zika will spread across the Americas

A number of tweets noted that Zika was reaching parts of the Americas:

'Zika looks like it is spreading very explosively across the Americas [URL]' (6.3.1)

'Looks like Zika is endemic in some parts of the Americas [URL]' (6.3.2)

In a similar manner to the previous theme, tweets within this theme could express a narrative of fear and expectation that the virus would rapidly spread across the world. The Americas is specifically mentioned because there were a number of Zika cases across this region (CDC, 2016).

6.4 Mentions of Brazil and Zika Virus

A number of Twitter users mentioned Brazil in the context of the Zika virus:

'Brazil not aware of the severity of the country's Zika virus [URL] (6.4.1)

'Brazil attempts to fight the spread of Zika' [URL] (6.4.2)

'Brazil has noted that the Zika outbreak is much worse than previously believed' (6.4.3)

Brazil was mentioned because of the incidence of Zika in the region (CDC, 2016). The articles also expressed a narrative that suggested Brazil may not have anticipated the severity of the Zika outbreak, as illustrated in *tweet 6.4.1* and *tweet 6.4.3.*

6.5 Geographical Tracking

A number of Twitter users tracked the geographical spread of the Zika outbreak:

'This is an interesting method of mapping the Zika outbreak [URL]' (6.5.1)

'Use our interactive explorer to track the Zika virus [URL]' (6.5.2)

Twitter users showed interest in tracking the incidence of the Zika outbreak; those inside Zika affected areas may have been more interested in tracking the spread of the Zika virus than those outside these areas. In the context of the Health Belief Model, assessing the location of the Zika virus may have allowed Twitter users to determine their perceived susceptibility to it.

6.6 Geographical Transmission

Twitter users expressed interest in the geographical transmission of Zika:

'Over 2 thousand are said to be affected by the mosquito-borne Zika in Colombia [URL]' (6.6.1)

'There are reports that a male in Australia may have contracted Zika [URL]' (6.6.2)

'Zika now in Costa Rica' (6.6.3)

'Zika now in New York' (6.6.4)

'There are fears that Zika could hit Britain' (6.6.5)

The Zika outbreak was an international threat reported across the world (Fauci and Morens, 2016). In the context of the Health Belief Model, Twitter users may have used the information related to the location of Zika to assess their perceived susceptibility to it.

6.7 Travel

A number of Twitter users were concerned about travelling during the Zika outbreak, which intersected with the Olympic Games taking place in Rio de Janeiro:

'Olympic games- because of Zika virus will cancel my trip' (2.2.1)

'64% of respondents from the USA noted that they will cancel travel to Zika-affected areas' (2.2.2)

These tweets highlight how the Zika outbreak caused concern to those who were planning to attend the Olympic Games. There were also a number of references to travel credits and refunds for those visiting Zika affected areas:

'Are you worried about Zika and travelling? Some airlines offering refunds [URL]' (2.2.3)

'Cruise liners are going to waive cancellation fees because of Zika' (2.2.4)

Citizens who were set to travel to Zika affected areas may have been concerned to do so because of the dangers associated with the disease. This may be particularly applicable to citizens who may have had a higher perceived risk severity of attaining Zika, such as pregnant women and couples seeking to start a family. The aviation industry was aware of the potential risk for passengers who were travelling whilst pregnant. British Airways, the largest UK airline, released the following statement (Telegraph, 2016):

> "If a pregnant customer is due to travel up to and including February 29, but they no longer wish to travel, they can change their booking free of charge, and delay their journey or amend to an alternative destination… this applies to flights to Brazil, Mexico, Barbados and the Dominican Republic."

The Health Belief Model is useful in this context because those who were most vulnerable to Zika, i.e. women who were pregnant, would be more likely to avoid travelling to areas affected by Zika due to the higher perceived severity of the virus.

7. General Discussions

Table 6-9 General Discussions

Theme	Sub-themes
7. General Discussions	**7.1 Zika Found in Ugandan Forest**
	7.2 Zika and Climate Change or Global Warming
	7.3 Broadcast Advert
	7.4 Information Seeking
	7.5 Politics
	7.6 Humour and Sarcasm
	7.7 Name Discussion
	7.8 Zika Information

7.1 Zika Found in Ugandan Forest

A number of tweets contained links to a news article that were referring to the origin of the Zika outbreak:

'The heart of Zika: The Forest Where Zika was located' (7.1.1)

'The country where Zika was first found hasn't been affected that badly' (7.1.2)

Zika may have been expected to be most prevalent in the region it was first discovered, i.e. Uganda. However, it appears that Zika had not affected the region as badly as other parts of the world, at dissonance to its Twitter visibility.

7.2 Zika and climate change / global warming

Certain narratives within the media linked the concept of climate change to the incidence of Zika:

'Climate change may be spreading diseases such as Zika [URL]' (7.2.1)

'Must know things about Zika and climate change!' [URL] (7.2.2)

The tweets within this theme may have arisen because some research has suggested a connection between extreme climate conditions and the emergence and transmission of Zika (Paz and Semenza, 2016).

7.3 Broadcast Advert

Twitter users linked to and shared broadcasts on TV and radio that on the topic of Zika:

'Professor Tom will be chatting to our station today at 1PM about Zika' (7.3.1)

'Be sure to follow our webinar on the Zika' (7.3.2)

There was a heightened interest in Zika during this time because it had been spreading across the world, and this was reflected in increased coverage by media and TV outlets. One method of promoting television programmes and radio broadcasts is by tweeting about them in order to reach larger audiences.

7.4 Information Seeking

A number of Twitter users sought information on Twitter during the outbreak:

'What is the Zika virus?' (7.4.1)

'Are the Olympic Games going to be safe as now there is Zika around – Is it worth the risk?' (7.4.2)

'How long should we wait until we reproduce? Until there is a Zika vaccine or until the virus vanishes?'(7.4.5)

The heightened interest around Zika may have led to public information needs on different aspects of the outbreak, as demonstrated by their tweets. The information from this theme may be of interest to health authorities who are in a position to disseminate health information and incorporate the answers to common queries.

7.5 Politics

A number of Twitter users referred to politics within their tweets:

'If Donald Trump was in charge I am sure there would be no Zika outbreak' (7.5.1)

'The PNP party can't handle an outbreak of Zika. 70% of people surveyed reported' (7.5.2)

Politicians and citizens may have used the Zika outbreak to score points against political rivals rather than express a genuine interest in the outbreak of Zika.

7.6 Humor and/or Sarcasm

A number of Twitter users shared tweets that were humorous and sarcastic:

'Olympics will act as the catalyst to spread the Zika faster around the world! ' (7.6.1)

'Was about to sleep then googled images of kids with Zika – OMG!' (7.6.2)

'Zika sounds like a new member of the spice girls' (7.6.3)

'Zika sounds like the name of a drag queen' (7.6.4)

Tweets within this theme were not offensive towards persons suffering from Zika *per se*, but would consist of general humorous and sarcastic views towards the Zika outbreak.

7.7 Name Discussion

Twitter users also discussed the name given to Zika:

'Zika is a disgusting name for a virus! ' (7.7.1)

'Folks might confuse Zika with Zica [URL] ' (7.7.2)

'The tata Zica car brings the Zika virus to mind' (7.2.3)

'Tata ponders over renaming the Zica car to Zika because of the spread of the Zika virus' (7.2.4)

Zika is an unusual word, and people may not have been aware of the term prior to the outbreak. There was also heightened interest around a car which was set to be named 'Zica' which led to the manufactures to rename it 'Tiago' in order to remove the potential link to the Zika outbreak (Weaver, 2016).

7.8 Zika Information

There were a number of Twitter users who shared useful information related to the Zika outbreak within their tweets:

'Here are some things that you must know about Zika [URL]' (7.8.1)

'The government has set up a group for monitoring Zika [URL]' (7.8.2)

'Zika requires a concerted effort to defeat [URL]' (7.8.3)

'This is how to stop the Zika virus [URL]' (7.8.4)

This theme shows that Twitter users shared informational tweets on a wide variety of topics which related to the Zika outbreak. One of the challenges of extracting insight from Twitter data is that it is not possible to drive Twitter users to tweet in a certain manner. This is also a strength, because it allows discussions to emerge that may not necessarily be in a survey or interview.

6.6 Frequency Distribution of Themes

Figure 6-2 below provides a visual overview of the different themes that were found to emerge and the size of the boxes reflect the frequency of occurrence.

Figure 6-2 Frequency distribution of themes

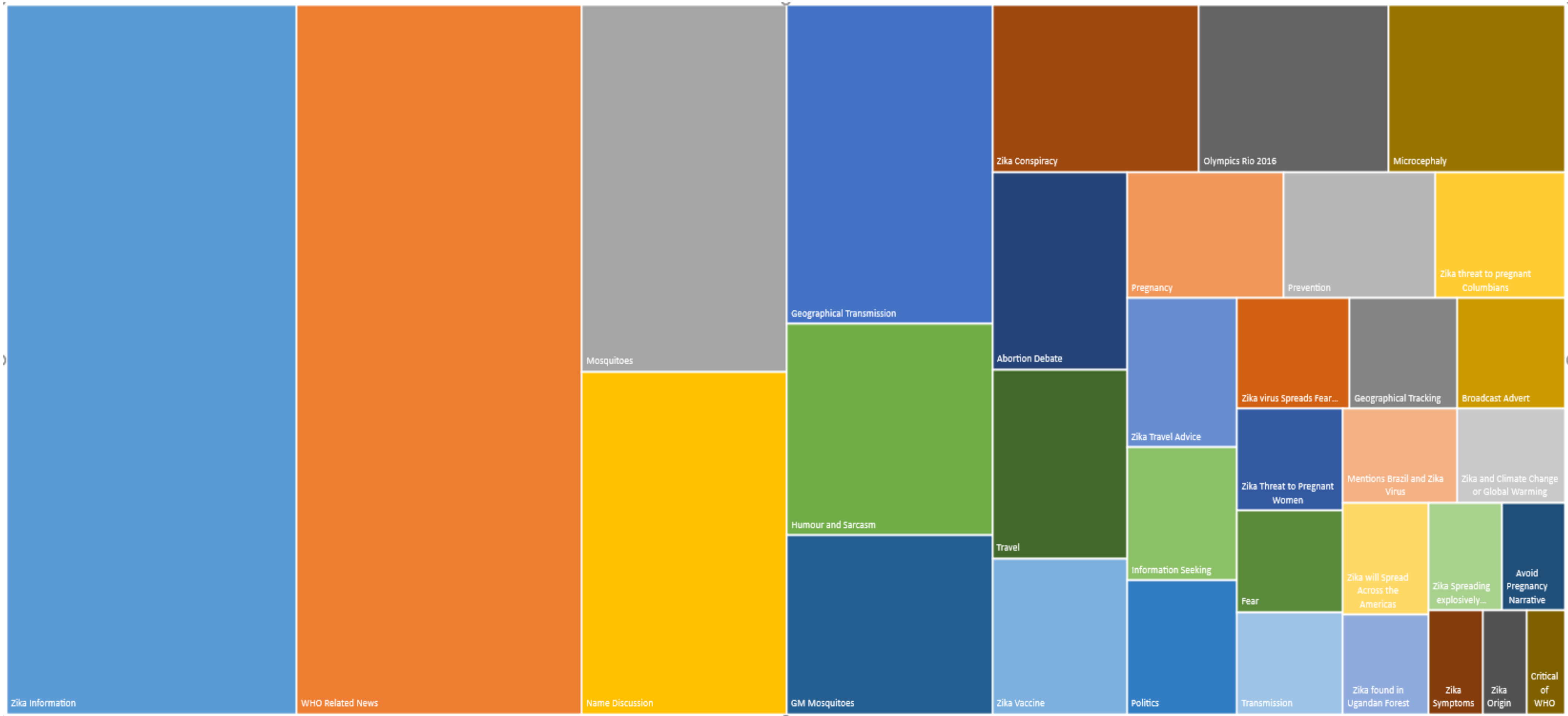
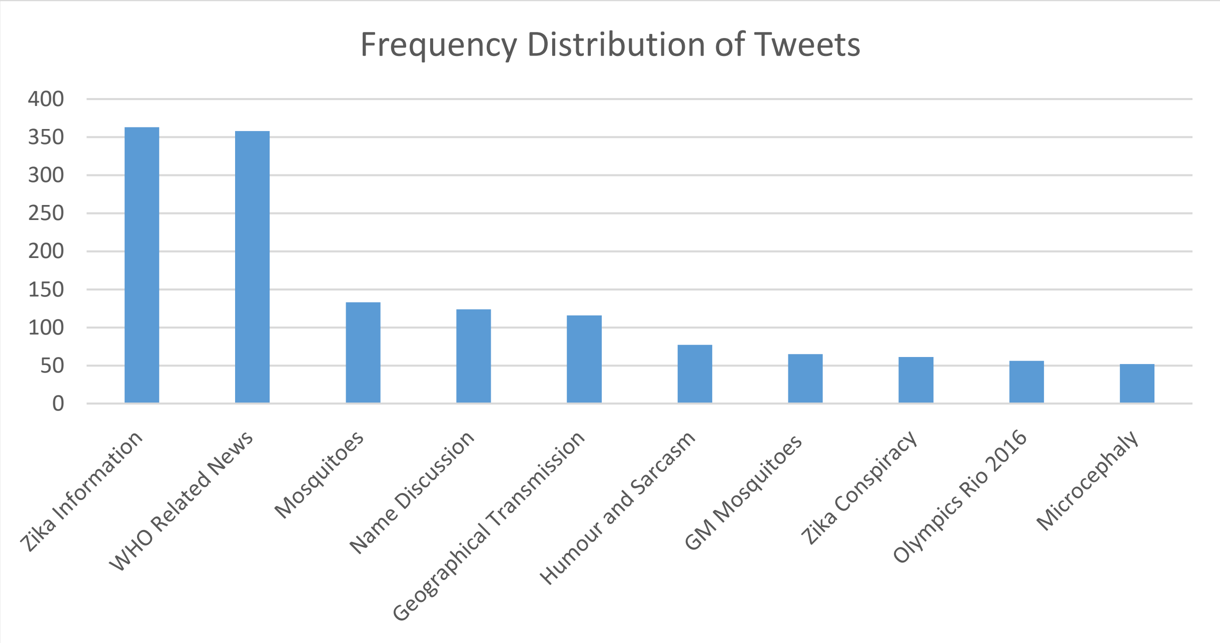


Figure 6-3 below is a bar-chart of themes which contained at least 50 tweets. It shows that the most popular tweets consisted of general Zika information (n = 363), WHO related news (n = 358), mosquitoes (n = 133), discussions around the name of Zika (n = 124), tweets which were humorous and/or sarcastic (n = 77), conspiracies around Zika (n = 61), the Olympics in Rio in 2016 (n = 56), and Microcephaly (n = 52).

Figure 6-3 Bar chart to show frequency distribution of tweets



Among the most frequently shared themes, it is interesting to note the theme of conspiracy theories appears. These results may be of interest to health authorities who are in a position to disseminate health information and who could assess the validity of conspiracy theories, sharing appropriate information.

6.7 Validity and Reliability of Results

6.7.1 Test Re-Rest Reliability

The principles of test-retest reliability were outlined in Methodology Chapter 3 section 3.14.2. In this study, the coder who coded the initial data (WA) re-coded a subset of data after a period of time had elapsed (three-months) in order to assess test-retest reliability. The percentage agreement was 99.37%, and κ =0.85, a substantial level (McHugh, 2012).

6.7.2 Intercoder Reliability

The concept of intercoder reliability was outlined in Methodology Chapter 3 section 3.10.5. In this study, an independent coder coded a subset of data. The intercoder reliability percentage agreement was 99.24% and κ =0.56, a moderate level (McHugh, 2012).

6.8 Evidence of Themes across the outbreak

Section 4.15 provides a justification for searching for themes across the outbreak period. In this study, Twitter's Advance Search was utilised in order to locate themes that had not been reported in previous literature and which could be identified by keywords. The themes were selected because they were not specific to the 2-day time period, were not reported in previous literature, and which were searchable by using keywords.

Table 6-10 Themes from across the outbreak

	Year 2016						
Themes	**January**	**February**	**March**	**April**	**May**	**June**	**July**
Name Discussion	x	x	x	x	X	x	x
Information Seeking	x	x	x	x	X	x	x
Olympics Rio 2016	x	x	x	x	X	x	x

The table above shows that certain themes such as the geographically tracking of Zika, humorous and sarcastic posts, discussions around its name, and tweets which were critical of the WHO would appear throughout the outbreak period.

6.9 Discussion

6.9.1 Key Findings and Implications of results

Overall, tweets contained much more content from news articles shared on the Zika than personal views. There were a number of varied discussions taking place around the Zika epidemic based on seven key discussions around pregnancy (164/8.36%), travel and the Olympics (120/6.12%), mosquitoes and conspiracy (259/13.21%), health organisations

(365/18.62%), health Information (160/8.16%), travel and tracking (246/12.55%), and general discussions around Zika (646/32.95%%). There was discussion surrounding pregnancy because Zika is known to be associated with a medical condition (microcephaly), which causes complications in children during birth (Mlakar, Korva, Tul, Popović, *et al.*, 2016). For women and couples in Zika affected areas, it was recommended that pregnancy be avoided. However, Twitter users highlighted the controversial nature of this recommendation in countries where access to contraception can be difficult to obtain. Moreover, some users also highlighted the difficulties of requesting an abortion on medical grounds because of tough laws surrounding abortion in some parts of Latin America. The Health Belief Model (HBM) is useful in this context because it indicates that people who deem themselves to have a high perceived susceptibility are more likely to undertake a particular action, i.e. that pregnant women would be more likely to avoid Zika affected areas. In the context of Zika, it is probable that the perception of benefits would outweigh barriers if travelling was non-essential because of the threat that Zika posed for pregnant women. This argument is further reinforced by the decision of airline companies to offer travel refunds and credits to women travelling in these areas.

Twitter users referred both generally and specifically to mosquitoes within their tweets. This included sharing conspiracy theories that genetically modified mosquitoes were responsible for the Zika outbreak, and that Zika had been manufactured. These findings may be of interest to public health organisations who could use this information to monitor further cases of Zika for these specific conspiracy theories and disseminate guidance and advice on their legitimacy.

Some Twitter users were also critical towards the WHO, or would share news originating from the WHO during this time period. The WHO may thus be interested to find that they were considered an authoritative information source during the outbreak period. The health information shared included information on the transmission of Zika and vaccines. Health authorities could utilise this information to disseminate information around the transmission of Zika and potential vaccines for the virus.

When reflecting on the number of general themes, there appeared a sense that the Zika virus was out of control: news articles would note that the virus was spreading ‘explosively’ in South and Central America and the Americas. Twitter users expressed an interest in geographically tracking the incidence of Zika. The information derived from tracking Zika could have allowed users to assess their perceived susceptibility of the virus.

A number of general discussions were also shared, which noted that Zika had been located in an Ugandan Forest, that it was linked to climate change, and expressed information on politics, and humorously /sarcastically. There were also tweets discussing the name given to Zika.

6.10 Limitations

The data selected was tweeted over a two-day period, and therefore some of the themes that emerged could be specific to this time period. However, section 6.8 did find examples of tweets from across the outbreak of Zika that conform to the determined themes. Moreover, if different time intervals were examined, it could be possible that different themes and sub-themes would emerge. Additionally, by examining English-language tweets the study is limited because it may not have captured the voices of those who were directly affected and/or in the front line in parts of Latin America where the Zika virus originally emerged. This is because Twitter users in these regions may have been tweeting in other languages such as Portuguese.

6.11 Summary

This chapter comprised of a smaller empirical study undertaken in order to examine the discussions taking place on Twitter during this time and to compare these to tweets on swine flu and Ebola. There were many news articles shared on Twitter during this time period, which formed a number of themes. The discussions attracting personal opinions and views centred on conspiracy theories and debates around abortion laws and ethics. This chapter has provided results from a smaller empirical study which analysed tweets on the Zika outbreak, and the next chapter discusses and compares the results of the individual cases to one another.

Chapter 7 Discussion

7.1 Introduction

The previous three chapters presented results from the thematic analysis of three case studies on Swine Flu (Chapter 4), Ebola (Chapter 5), and Zika (Chapter 6) from a time when there was a heightened interest in the outbreaks. This chapter compares and contrasts the similarities and differences of the cases. Section 6.2 provides the background for this chapter, section 7.3 compares themes identified to one another, section 7.4 provides guidance on disseminating information during an infectious disease outbreak. Section 7.5 reflects on the utility of the Health Belief Model, and section 7.6 discussed the applicability of Information theory. The chapter then draws on literature from other fields and considered the results of the study under the concept of the moral panic (section 7.7), and the outbreak narrative (section 7.8). The chapter then considered the effectiveness of Twitter for qualitative research (section 7.9).

7.2 Background

A comparison between the two cases of Ebola and swine flu is important because there might be discussions taking place on social media that are common to infectious disease outbreaks, particularly when there is an increased interest in them. The comparison will also draw parallels the case of Zika. It is important to note that this present study is the first to conduct an in-depth analysis of tweets on three different infectious disease outbreaks. This comparison will help create guidelines that provide advice on the types of information Twitter users may require during an infectious disease outbreak and emergencies. On Twitter, discussions may emerge that are not necessarily included in a survey and/or interview questions.

7.3 Comparison

Sections 7.3.1 to 7.3.7 compare the results of the themes from the individual cases to one another, section 7.3.8 provides an overview of the similarities identified in themes. Section 7.3.9 will compare the results of themes for swine flu to previous research, section 7.3.10 provides a comparison to previous research for Ebola, and section 7.3.1 for Zika.

7.3.1 **Emotion and Feeling**

There was overlap across the two cases for the theme of *emotion and feeling*. For instance, it can be seen that both cases resulted in Twitter users expressing general fear towards each of the outbreaks. There were similar sentiments shared in both cases because Twitter users would mention that they were afraid of the diseases, as in the examples below:

'Swine Flu is scary' (A1.1)

'Wow, Ebola is scary' (E2.1)

In the case of Zika, Twitter users would also express their fear of Zika:

'Zika is here – God help us' (2.3.2)

For Ebola, some Twitter users mentioned that they had fears related to Ebola mutating and finding its way to the US, as illustrated in the examples below:

'If Ebola comes to the U.S. I think we are finished, so scared' (E2.3)

'Very fearful of Ebola mutating, think we should help the affected' (E2.4)

The fears above were specific to the Ebola outbreak. The comparison highlights how both cases had a specific sub-theme entitled "*fear of travel*". There were similarities in both cases where Twitter users were unsure of whether they should travel internationally. Illustrative tweets from each of the cases are provided below:

'We thought of visiting Morocco, but are unsure now if it is safe for the children, don't want to get Ebola!' (E3.1)

'Trip to Mexico is next week – but now wondering what shall I do?' (A2.8)

Both the Ebola and swine flu outbreaks appear to have caused Twitter users to develop specific travel-related fears. A potential reason for this is that both diseases were perceived to be deadly, and travelling may have increased the perceived risk of catching them. This consideration may have implications for the hospitality and tourism sector because there may be decreased interest in travelling, if there were to be an international health scare. In a study published in *The Lancet*, it was noted that air travel via commercial airlines is an apt environment for the spread of infectious disease, which can, in fact, occur more than officially reported, due to a bias in reporting from the airline companies (Mangili and Gendreau, 2005). The Health Belief Model is useful in this context because, if Twitter users were travelling, they may have felt that their perceived susceptibility to either swine flu or Ebola would increase. When considering the Ebola case study, a specific sub-theme entitled '*Dead rising generates fear'* was found to emerge, which was related to a news article which suggested that Ebola patients were rising from the dead (See Section 5.6). For Ebola, another specific theme that emerged was that of users expressing prayer, and this occurred in 0.6% of the Ebola sample of tweets. Some Twitter users had family in West Africa. Modern healthcare facilities do exist in Africa; however, a number of communities in West Africa may still rely on traditional and spiritual healing, and this could explain why some of these tweets were sent. During the 2014 Ebola outbreak, there were reports of a number of cases where faith healers would pray for Ebola victims, and would end up contracting the disease and spreading it to other individuals (Manguvo & Mafuvadze, 2015). Manguvo & Mafuvadze (2015) noted that awareness programs should be launched which would target geographical areas that are at a high risk for traditional and spiritual healing. Twitter could be leveraged in order to disseminate this information, as research has found Twitter is a widely utilised platform in West Africa (Karanja and Flanagan, 2012).

7.3.2 Health Information

There were similarities in the *health information* theme for swine flu, Ebola and Zika. It was found that there was overlap with the theme *symptoms*. In both Ebola and swine flu, when Twitter users were facing minor ailments they would question and/or state whether they had caught the disease, as illustrated in the examples below:

'I have a headache: swine Flu or normal flu?' (B5.1)

'My throat is hurting; I think I have Ebola' (F3.1)

These findings can be of interest to the public health sector because they indicate that, at the time of an outbreak, members of the public will want and need information that allows them to differentiate between symptoms associated with non-life threatening illnesses, such as the common flu, and more serious conditions, such as swine flu. Failing to do so could cause hospital and emergency visits to increase. Therefore, this information should be disseminated as a matter of high priority during an infectious disease outbreak. In both cases, Twitter users were interested in mapping the geographical spread of the conditions in order to assess the proximity of a disease to their location, as illustrated in the examples below:

'Just checked and Swine Flu has been reported 40 miles from north of Austin, that's getting mighty close to us!' (B2.1)

'There is Ebola patient in DC? It is getting close' (F1.2)

In the case of Zika, Twitter users would also display interest in tracking the incidence of the virus:

'Zika now in Costa Rica' (6.6.3)

Twitter users might have wanted to track the spread of an infectious disease in order to feel safe and maintain a form of control. Additionally, Twitter users may have wanted to know how a disease can be treated because, in the case of Ebola, there was discussion related to vaccines and, in the case of swine flu, Twitter users would discuss medications, as illustrated in the examples below:

'Tamiflu can protect against Swine Flu' (B7.1)

'Why is there no vaccine for Ebola yet?' (F4.2)

Twitter users shared Zika news articles on Twitter related to vaccines:

'Everyone around the world should now be aware of Zika, and I really hope there is a vaccine soon' (5.5.2)

In these cases, it was also found that there were a small number of Twitter users who were speculatively diagnosing themselves. For swine flu, Twitter users would talk about diagnosis in general terms. Other Twitter users discussed prevention. In the case of swine flu, there was also a specific sub-theme entitled *prevention products* which occurred in tweets. Illustrative tweets are provided below:

'Come on people! Wash your hands #swineFlu!' (B3.3)

'Wash your hands or you can spread Ebola' (F5.1)

Twitter users also discussed the transmission of the diseases. Illustrative tweets from each of the cases are provided below:

'Human to human contact can spread Swine Flu' (B1.3)

'To the best of my knowledge water does not spread Ebola' (F2.3)

'Mass gatherings have the potential to begin outbreaks like Zika' (5.3.1)

In the case of Ebola, many tweets surrounded the concept of quarantine. For swine flu, Twitter users drew parallels with previous outbreaks, whereas with Ebola these comparisons were not made. These differences in themes could have occurred by chance, i.e. that on the days the data were sampled, people mentioning Ebola referred to quarantine and people mentioning swine flu referred to other diseases. Other possibilities are that users might have been more familiar with the swine flu outbreak, since it had occurred previously in the West. Twitter can act as a platform for health organisations to disseminate health information during public health emergencies. Importantly, in this case, Twitter was also used by citizens to share health information, for example on prevention and good personal hygiene.

7.3.3 General Commentary

These categories contained a high number of tweets because, in both case studies, there were many tweets which were not insightful and difficult to code into themes and sub-themes in their own right, as illustrated in the examples below:

'Hurray! My 500th tweet is on Swine Flu!' (C1.3)

'Just found out about the Ebola epidemic' (H1.2)

In the Zika case, there were also a number of miscellaneous tweets. Previous research has also found there to be many non-relevant tweets when analysing Twitter data using qualitative methods (Chew and Eysenbach, 2010). Some authors may choose to remove general discussions completely when reporting results. For example, Hewis *et al*. (2015) devised an inclusion criterion based on whether the tweet matched the topic under study. As mentioned throughout this thesis, this was one of the challenges identified when using Twitter as a source of data for in-depth qualitative insights, because in a survey or an interview non-relevant content can be kept to a minimum.

In both cases Twitter users would seek information and illustrative examples from the two cases are provided below:

'What is Ebola?' (H2.1)

'How do you know if you have Swine Flu?' (C2.1)

In the case of Zika Twitter users would also seek information:

'Are the Olympic Games going to be safe as now there is Zika around – Is it worth the risk?' (7.4.2)

These findings suggest that members of the public who use Twitter during the peak of an infectious disease outbreak may do so to seek information from other users. An interesting difference observed with the Ebola case is that Twitter users would share links to Instagram (an image sharing platform) in tweets. During the swine flu outbreak in 2009, Instagram was not available, therefore there were no tweets linking to it. In future epidemics and pandemics, health authorities could monitor whether Twitter users link to new social media platforms, and the content on these platforms should be better understood.

Another interesting finding from the comparison is that Twitter users were sharing links to YouTube for the Ebola outbreak. Therefore, media such as images and videos may have played a greater role in the outbreak of Ebola, as Twitter users linked to external video and image sharing platforms. With the earlier swine flu case study, there may have been fewer platforms available for Twitter users to provide Web links to. For Ebola, Twitter users linked to other tweets, while this practice was not observed with the swine flu dataset. In both cases, Twitter users would share content related to the economic impact that the outbreaks were having on society, illustrated in the examples below:

'Stock market is not looking good due to Swine Flu' (C3.2)

'The price of cocoa has fallen dramatically because of Ebola [URL]' (H3.1)

In both cases there were Twitter users who were downplaying the outbreaks:

'The regular Flu by itself kills so many people – people need to relax #swineFlu' (C4.2)

'There is not really a chance of an Ebola outbreak in America' (H10.1)

For Ebola, there was a specific sub-theme related to the sharing of conspiracy theories. There were themes that appear to be specific to the individual cases. For example, the theme entitled *western privilege* appeared to be specific to the outbreak of Ebola, while the sub-theme *name discussion* was specific to the swine flu outbreak and surrounded discussions on the term 'swine'. Overall, there appeared to be very little in common across this theme because Twitter users focused on different aspects of the outbreak and offer their views towards it.

7.3.4 Official Organisations

When discussing official organisations in the case on swine flu, Twitter users would either mention media or health organisations in their own right and/or offer a personal opinion of them. For Ebola, the mentions of health organisations would derive from news articles that Twitter users were sharing on the platform. In the case of the swine flu outbreak, users were critical of health and media organisations. The themes were not directly comparable because for swine flu the mentions included personal opinions whereas the mentions for Ebola were based on content that would express general views. In the outbreak of the Zika, there were Twitter users who were critical of the WHO and share news articles which contained information from the organisation. However, in all cases there were some references to the CDC and the WHO, which could indicate that these organisations will be influential in future infectious disease outbreaks.

7.3.5 Politics

Obama was mentioned frequently in both Ebola and swine flu case studies, and many users blamed him for the each of the outbreaks:

'Wonder if there is a correlation between Obama's PIG filled stimulus and outbreak of Swine Flu?' (E2.1)

'Our so-called president Obama still has not come out with a plan to address Ebola' (K1.3)

In the political sphere, opponents may use current events to criticise the party in government, and citizens may also engage in this. It could be, therefore, that members of the Republican Party and citizens who were generally against Obama had attempted to use the swine flu and Ebola outbreaks against him**.** In the Ebola case study, there were specific sub-themes related to Julie Bishop, the Australian Prime Minister at that time, who was openly criticised. In the Zika case, the U.S. president at the time, Donald Trump, was also referenced:

'If Donald Trump was in charge I am sure there would be no Zika outbreak' (7.5.1)

Research has examined the relationships between new digitised forms of political activism in association with the rise of social media (Gerbaudo, 2012). For example, Loader and Mercea (2011) noted that social media platforms might have led to a more participatory political culture where users of social media platforms would offer their opinions towards emerging political events. Moreover, they noted that opinions from social media users, who may be non-political-experts, may be relayed by the media and would have the potential to over-shadow expert commentary. This is particularly relevant for present study because there were mainstream media news articles written specifically based on tweets sent to criticise President Obama (USA Today, 2017).

7.3.6 References to Geographical Location

Differences were seen in the geographical locations mentioned by Twitter users in these case studies, because of the geographical locations of the outbreaks themselves. For the swine flu case study, there were tweets that referred to Mexicans, borders, and there were many references to Mexico or Mexico City, because this is where the swine flu outbreak originated. For Ebola, the locations of interest related to Sierra Leone, Liberia, Nigeria, and Guinea. In the case of Zika, the geographical areas mentioned were Brazil, South and Central America, and the Americas. There were also references to Uganda, which is where the virus is said to have originated. During the Zika outbreak, the Olympic games were taking place in Brazil, which is why the specific area was mentioned. The geographical areas affected might also be subject to stigma and/or negative sentiments from some Twitter users, as observed in the tweets related to swine flu. Ostherr (2005) also highlighted that not being able to track the location of a disease might raise the anxiety levels of the general public. In some media narratives, being able to track a disease supposedly minimises its risk, and delivers a sense of control. This might explain why there was an increased interest around tracking the location of the swine flu and Ebola outbreaks. Therefore, in future infectious disease outbreaks, there might be an increased interest surrounding the geographical locations where an outbreak is reported.

7.3.7 Sarcasm and Humour

In the case studies Twitter users shared tweets which were humorous and/or which used sarcasm. Illustrative tweets from each of the cases are provided below:

'Had a quick scan of the news – interesting how everyone they are showing with Swine Flu is ugly!' (H4.3)

'What meal would you like to cook with your ex...Ebola' (L2.2)

Sarcasm also occurred in all three case studies. Illustrative tweets from each of the cases are provided below:

'My squad has the Ebola' (L1.6)

Omg, Swine Flu is spreading on Twitter! (H5.3)

'Was about to sleep then googled images of kids with Zika – OMG!' (7.6.2)

A key difference between the two cases was that, in the swine flu case study, Twitter users were specifically humorous towards pigs, leading to a specific sub-theme *humour related to pigs*.

Twitter users were not humorous towards the outbreak situation of swine flu, Zika and Ebola *per se,* but were using the names of the diseases in a context that would be humorous. There were also similarities in the extent to which users would refer to zombies and the potential of a 'zombie apocalypse'. For swine flu, this sub-theme was named *popular culture/understanding* and for Ebola there were the sub-themes entitled *zombies* and *zombie apocalypse*. These were different because whilst Ebola tweets would specifically mention zombies and the potential for a zombie apocalypse, swine flu tweets would refer to a wide range of topics from across popular culture.

Illustrations of these tweets which made reference to zombies across the two cases are provided below:

'I am so ready for when the Swine Flu starts to turn people into flesh eating zombies' (H3.4)

'Did anyone else hear that there might be zombies in West Africa?' (L3.2)

It is interesting to note that, in both swine flu and Ebola cases, which were five years apart, Twitter users would refer to zombies in the context of the outbreaks. It might be that the narratives of infectious disease outbreaks lead Twitter users to draw comparisons with fictionalised cases of zombie outbreak scenarios. The emergence of sarcasm and humour themes could indicate that Twitter users may not perceive infectious disease outbreaks as serious. However, as was noted in the case study related to Ebola (Chapter 5), linguistic researchers argue that humour on social media platforms could be a way for users to show solidarity (Zappavigna, 2012). Hence, one explanation for why there were so many tweets expressing sarcasm and humour was that, in some instances, this was a way for users to come together and show solidarity collectively. Research has also found that humour has the potential to be beneficial psychologically (Heath & Blonder, 2003; Chapple & Ziebland, 2004). Zsófia (2016), who investigated humour in relation to online discussions surrounding cancer, noted that “humour helps shift perspective on a stressful situation, and through reappraisal from a less threatening point of view, the stress becomes more manageable” (p.28). Therefore, it could be that humour was used by Twitter users in order to manage stress: when something is depicted as silly and humorous, it is less likely to come across as scary or life-threatening. These discussions might have allowed Twitter users to form a sense of community on Twitter, and collectively take part in sharing humorous content. However, this is not likely to be the case with all the tweets that were of a sarcastic and humorous nature.

7.3.8 Overview of Similarities and Differences

Overall, this comparison exercise has shown that there are indeed several similarities and differences in the way in which Twitter users responded during the 2009 swine flu pandemic the 2014 Ebola epidemic and the Zika virus outbreak. Figure 7-1 provides an overview of the similarities and differences that were found in each of the cases.

Figure 7-1 Overview of similarities and differences between Ebola and swine flu

Similarities and Differences between cases

Similarities:

- **Fear**
- **Anger**
- **Fear of travel**
- **Transmission**
- **Prevalence monitoring**
- **Speculative diagnosis**
- **Prevention**
- **Symptoms**
- **Medications e.g. vaccines**
- **Economic impact of disease**
- **Information seeking**
- **Voice of reason or downplaying the outbreak**
- **General discussions**
- **References to official organisations**
- **References to Obama**
- **References to areas affected**
- **Humour**
- **Sarcasm**
- **References to popular culture**

Differences for Ebola:

- **Dead rising generated fear**
- **Praying, Prayer or call to God**
- **Quarantine**
- **Link to Instagram**
- **Links to YouTube**
- **Links to other tweets**
- **Conspiracy Theories**

Differences for Swine Flu:

- **Worry**
- **Prevention Products**
- **References to other infection or disease**
- **Unfollowing users**
- **Frightening Scenarios**
- **Images used in tweets**
- **Name Discussion**
- **Humor related to pigs**

Figure 7.1 shows that there were similarities across at least 19 themes and the outbreaks would evoke a similar response from Twitter users. Some of the differences in Figure 7-1 may have been due to the characteristics of the diseases and to specific events. For example, with swine flu there were discussions surrounding its name and the potential to cause confusion. With Ebola, the story that people who had died had come back to life was specific to events that occurred during that outbreak. Other differences related to the availability and popularity of video sharing and image sharing platforms such as Instagram and YouTube.

The similarities for Zika primarily relate to the theme of fear, mentions of the WHO, political references, transmission, prevention, and travel. Twitter users also discussed the name of Zika, as was the case with swine flu. However, discussions around Zika were significantly distinct to Ebola and swine flu, and there were a number of interesting themes specific to the Zika outbreak were as followed:

- Avoid Pregnancy Narrative
- Zika Threat to Pregnant Women
- Zika virus Spreads Fear Among Pregnant Brazilians
- Zika threat to pregnant Columbians
- Abortion Debate
- Pregnancy
- Olympics Rio 2016
- Mosquitoes
- Mosquitoes
- Microcephaly
- Zika Spreading explosively in South and Central America
- Zika will Spread Across the Americas
- Mentions Brazil and Zika Virus

Many of these differences in themes arose because the content within the tweets were related to the specific events and debates surrounding the Zika outbreak.

7.3.9 Comparisons with previous research (swine flu)

There are several studies which have examined Twitter content on swine flu (Chew and Eysenbach, 2010; Signorini, Segre and Polgreen, 2011; Kostkova, Szomszor and St. Louis, 2014). However, only one of these studies, i.e. Chew and Eysenbach (2010), examined the content surrounding swine flu by using content analysis and placing tweets into themes and sub-themes. In that study, a total of 5,395 tweets were examined from May 11th to December 21st 2009 using the keywords ‘H1N1’ and ‘swine flu’. This present study utilised a different method to sample Twitter data. Tables 7-1 and 7-2 below show the results of categorisation of tweets from this study.

Table 7-1 Summary of key themes that were identified by Chew and Eysenbach (2010)

Theme	Description	Total
Resources	Tweets with news general updates or information.	2840 (53%)
Personal Experiences	A personal direct or indirect experience.	1214 (23%)
Personal Opinion	User provides their opinion on the outbreak.	740 (14%)
Jokes	Tweet has a H1N1 joke.	421 (8%)
Marketing	Advertisements related to products on H1H1.	72 (1%)
Spam	Unrelated tweet.	108 (2%)

Table recreated from Chew and Eysenbach (2010)

Table 7-2 Summary of key qualifier themes that were identified by Chew and Eysenbach (2010)

Sub-Theme	Description	Total
Humour	Tweet is sarcastic or humorous.	687 (13%)
Relief	Tweet notes happiness, expressed joy or a sense of peace.	81 (2%)
Downplayed Risk	Tweet aims to de-emphasise the outbreak and to bring it in perspective. It may also show a lack of interest.	106 (2%)
Concern	Tweet with a H1N1 worry, fear, anxiety, or feelings of sadness for other people.	633 (12%)
Frustration	Tweet expressed scorn, anger, annoyance. These types of tweets might also include coarse language.	212 (4%)
Misinformation	Tweet contains misinformation and will include conspiracy and/or doomsday theories.	243 (5%)
Question	Tweet will ask a question and/or contain a question mark.	555 (10%)

Chew and Eysenbach (2010) found that, based on their analysis, the theme with the largest percentage of tweets was *resources* (53%). This is similar to the theme of *general discussions* (31.8%) found in the qualitative analysis of swine flu tweets (Chapter 4). This highlights a difficultly of using Twitter data to gain insights into public views and opinions on specific topics as there are likely to be off-topic discussions taking place. Other similarities can be found with the theme of humour, downplayed risk, voice of reason, question (i.e. information seeking).

Table 7-3, below, shows how the themes from this present study fit the thematic headings from Chew and Eysenbach (2010).

Table 7-3 Similarities between present study and themes from Chew and Eysenbach (2010)

Main Theme from Chew and Eysenbach (2010)	**Themes from this present study**
Resources	C.1 General Discussions (1826/31.8%) C.2 Information Seeking (145/2.5%) C.3 Economic Impact of Swine Flu (62/1.1%) C.4 Voice of Reason (109/1.9%) C.5 Frightening Scenarios (13/0.2%) C.6 Name Discussion (26/0.5%) C.7 Resources (42/0.7%) C.8 Images used in Tweets (36/0.6%) C.9 Unfollowing Users (2/0.03) C.10 Other Discussions (206/3.6%) D.1 Health Organisations (general) (136/2.4%) D.3 Media Organisations (general) (444/7.7%) E.2 Obama (43/0.75%) A.3 Anger (17/0.3%)
Personal Experiences	A.2 Fear of Travel (54/0.9%) C.5 Frightening Scenarios (13/0.2%) B.6 Speculative Diagnosis (18/0.3%)
Personal Opinion	E.1 Political Reference (81/1.4%) E.2 Obama (43/0.75%) D.2 Health Organisations (critical) (7/0.1%) D.4 Media Organisations (critical) (88/1.5%) F.1 Reference to Mexico and/or Mexico City (162/2.8%) F.2 Reference to Mexicans (43/0.8%) F.3 Reference to Borders (6/0.10%)
Jokes	H.1 Humour Related to Pigs (100/1.8%) H.2 Nervous Humour (18/0.3%) H.3 Popular Culture/Understanding (221/3.9%) H.4 Miscellaneous Humour (378/6.6%) H.5 Sarcasm (258/4.5%) G.2 Food Humour (92/1.6%)
Marketing	B.4 Prevention Products (126/2.2%)
Spam	I.1 Not Relevant (1,937/33.7%)

Table 7-3 shows how the themes that emerged from this present study on swine flu relate to the themes from Chew and Eysenbach (2010). Additional themes not found in Chew and

Eysenbach (2010), but which have emerged from the present study are highlighted figure 7-2 below.

Figure 7-2 Newly emerged themes for swine flu

Newly Emerged Themes for Swine Flu

- **Fear of Travel**
- **Prevalence Monitoring**
- **Prevention Techniques**
- **Prevention Products**
- **Symptoms**
- **Medication**
- **References to Other Infection or Disease**
- **Economic Impact of Swine Flu**
- **Frightening Scenarios**
- **Name Discussion**
- **Unfollowing User**

- **Health Organisations (critical)**
- **Media Organisations (critical)**
- **Political Reference**
- **Obama**
- **Reference to Mexico and/or Mexico City**
- **Reference to Mexicans**
- **Reference to Borders**
- **Pork Consumption**
- **Food Humour**
- **Humour Related to Pigs**
- **Popular Culture/Understanding**
- **Images used in Tweets**

The table above shows that there were a number of themes that have emerged in this present study which have not been reported in previous research. It is surprising to find that no previous evidence-based research uncovered discussions around the name that was given to swine flu, as well as whether it was safe to consume pork. In order to undertake thematic analysis correctly, a requirement is for all tweets within a dataset to be read and labelled, whereas with content analysis a code frame can be created by exploring a subset of tweets. This means that, with thematic analysis, important themes can develop at any stage of the coding process, and it appears that a significant amount of new themes has emerged in this way.

7.3.10 Comparisons with Previous Research (Ebola)

Studies on Ebola have looked at other aspects such as trends of information spread (Odlum and Yoon, 2015), analysis of a live Twitter chat (Lazard, Scheinfeld, Bernhardt, Wilcox, and Suran, 2015), and rumour detection (Oluwafemi, Elia, and Rolf, 2014) rather than providing insights into the discussions taking place. However, in only one of such studies the content of discussions was examined. Jin *et al.* (2014) conducted a computer-assisted content analysis of Twitter discussions on Ebola from September to October 2014 and identified four main clusters related to risk factors, prevention education, disease trends, and compassion as shown in Table 7-4.

Table 7-4 Clusters and keywords from Jin *et al*. (2014)

Clusters	Keywords
Risk Factors	Transmitted, Spreads, Human, Africa, Contact with, Contact, Animals, Infected, Through, Wild animals
Prevention Education	Signs Symptoms, Pigin, In Pigin English, And Preventative Measures Yoruba, Measures in Pigin, English Pigin English, Symptoms and Preventive
Disease Trends	Virus, WHO, You need, US, Africa, Need to know, Spread, Need to West, you
Compassion	Pray, Liberia, Help us, Sierra Leone, Guinea, Pray for, Nigeria, US, Guinea Liveria, Pray for Sierra, Pray

There are similarities in the themes located in the study above to this present study such as transmission, symptoms, WHO, Praying, Liberia, Nigeria, and Sierra Leone. Table 7-5 below shows the similarities between the themes in this present study to the clusters identified by Jin *et al*., (2014).

Table 7-5 Similarities between present study and clusters and keywords from Jin *et al.* (2014)

Clusters	Themes in support
Risk Factors	F.2 Transmission of Ebola (26/0.60%)
Prevention Education	F.5 Prevention (22/0.51%) F.3 Symptoms (37/0.90%)
Disease Trends	I.8 WHO (17/0.40 %) F.1 Transmission Reporting (41/1.00%)
Compassion	I. 1 Sierra Leone (104/2.42%) I.2 Liberia (36/0.84%) I.3 Nigeria (33/0.80%) I.4 Guinea (8/0.2%)

A number of themes and sub-themes identified only in this present study as shown Figure 7-3 below.

Figure 7-3 Newly emerged themes for Ebola from this present study

Newly Emerged Themes For Ebola

- **Anger**
- **Fear**
- **Fear of travel**
- **Dead rising generates fear**
- **Vaccines**
- **Prevention**
- **Speculative Diagnosis**
- **Quarantine**
- **Ebola patients rise from dead**
- **Australia will not send volunteers**
- **U. S to send troops to fight Ebola**
- **News story uses terrorism analogy**
- **Doctor exposed to Ebola**
- **FDA warning over fake drugs**
- **General discussions**
- **Information Seeking**
- **Economic Impact of Ebola**
- **Death Count**
- **Western Privilege**
- **Twitter users linking to other tweets**

- **Downplaying Ebola risk**
- **Conspiracy Theories**
- **CDC**
- **MSF**
- **UNICEF**
- **Sierra Leone**
- **Liberia**
- **Nigeria**
- **Guinea**
- **Obama**
- **Julie Bishop**
- **Critical of Governments**
- **Sarcasm**
- **Humour**
- **Zombies**
- **Zombie Apocalypse**
- **Ebola used as an insult**
- **Link to Instagram**
- **Twitter users linking to YouTube**
- **Refers to iPhone**
- **Twitter users linking to other tweets**

For Ebola, as shown in the figure above, a number of new themes were found to emerge which had not been reported in previous literature. Thus, this study has shed new light on the types of discussions that were taking place during the peak of the 2014 Ebola outbreak.

7.3.11 Comparisons with Previous Research (Zika)

Previous research on Zika by Fu *et al.* (2016) has found there to exist five key themes: the impact the outbreak was having on society, responses to Zika such as from the general public, government and private sector, pregnancy and microcephaly, the transmission of Zika, and a theme with case reports of Zika. Stefanidis *et al.* (2017) found people would be interested in tracking the Zika virus because it would spread across the world. They also found that tweets frequently mentioning pregnancy and abortion. The following themes appear to have not been reported in previous literature:

- Name Discussion (124/6.32%)
- Information Seeking (26/1.32%)
- Humour and Sarcasm (77/3.92%)
- Olympics Rio 2016 (56/2.85%)

7.4 Information Dissemination Guidance

After comparing the results of swine flu to those of Ebola, similarities in the content were identified that could be of interest to users. The table below was created by reflecting on these similarities, and by drawing on previous qualitative studies (Chew and Eysenbach, 2010; Jin *et al.*, 2014).

Table 7-6 Information dissemination guidance

Arises from Theme(s)	Context	How to address
Fear Anger	There may be Twitter users who feel fearful and angry during an outbreak.	Information on stress reduction strategies could be shared especially during peak time such as an unfolding event
Travel	There may be Twitter users concerned about whether they should travel abroad.	Travel updates (as well as actual risks in the area) could be shared, especially those regarding affected areas.
Transmission	There may be Twitter users who have information needs around the transmission of a virus.	Information on how a disease is transmitted could be shared.
Prevalence Monitoring	Twitter users may want to know about the geographical spread of a disease in relation to them.	Platforms for tracking the geographical spread of a disease could be tweeted.
Speculative diagnosis Symptoms	Twitter users may want to know how best to identify symptoms of a disease.	Information on how to identify symptoms could be disseminated on Twitter.
Prevention	Twitter users may like to know how to best avoid catching an infectious disease.	Best tips for prevention could be shared, and Twitter users could be promoted to retweet this information.
Medications, i.e. vaccines	Twitter users may wish to know whether there are medications available that could treat a disease.	Updates on preventive medications could be provided, i.e. their availability, and cost.

Table 7-6 above could be utilised in a future infectious disease outbreak by health authorities who could consider incorporating some of the information in their social media content sharing plans. However, it is also possible that the recommendations set out in the table may already be in use by health authorities and organisations. However, the suggestions may still be useful as these have been developed directly from analysing tweets and it may be useful to compare these guidelines to those already in use by health authorities. A novel aspect of the recommendations developed in this thesis, therefore, are that they have emerged directly from the analysis of Twitter data.

7.5 Utility of the Health Belief Model

The Health Belief Model was useful in interpreting the results during the analysis of Ebola and swine flu tweets. The constructs were applied during the writing of the thematic results, and the table below summarises where the model was applied and the context of its application, alongside an explanation for its use.

Table 7-7 Themes where the Health Belief Model was applied across Ebola and swine flu tweets

		Health Belief Model		
Theme	Context	Individual Perception	Modifying Factors	Likelihood of Action
Transmission	There was an increased awareness of potential behaviours and symptoms related to transmission. Because the perceived severity is high, the heightened media coverage around infectious disease outbreaks may increase the perceived threat of a disease.	Perceived Severity	Perceived Threat	N/A
Prevalence Monitoring	Twitter users were using Internet tools to track the location of swine flu. Users may have been engaging with this behaviour to assess the severity of the outbreak.	Perceived Severity	Perceived Threat	Some users may be more likely to engage in behaviours associated with tracking a disease.

Prevention techniques	Twitter users were angry at those who were seen to be going against the norm (e.g. covering their mouths when coughing or sneezing). People's perceived susceptibility would increase if people saw others who going against recommended behaviours. Moreover, users may be more likely to engage in recommended behaviour because if perceived severity of the disease is high as they are more likely to undertake the recommended behaviour.	Perceived Severity	Perceived Threat Cues to Action	More likely to engage in behaviours associated with prevention techniques.
Prevention Products	Twitter users noted that facemasks were not effective in protecting against swine flu. If people believe that engaging in new behaviours is not going to reduce their susceptibility of a disease then they are less likely to engage in such behaviours.	Perceived Susceptibility	Perceived Threat Cues to Action	Less likely to engage in purchasing prevention products such as facemasks.
Reference to other infection or disease	Users would reflect on previous outbreaks, and might be more likely to engage in certain behaviours if they feel other deadly diseases were also harmful.	Perceived Severity	Perceived Threat Cues to Action	More likely to engage with prevention behaviours.
Health Organisations General	An increase in the WHO alert level was cause for alarm and fear. Increasing the alert level by WHO may increase how users perceive the threat of a disease.	Perceived Severity	Perceived Threat Cues to Action	

Pork Consumption	Twitter users stopped eating pork or indicated they felt afraid after consuming pork products. Users may have thought consuming pork would increase their perceived risk.	Perceived Severity	Socioeconomic Knowledge Perceived Threat Cues to Action	Users might begin to cease the consumption of pork products.
Fear	Fear and specific fears related to the potential of a virus mutating. The perceived susceptibility of a disease may be high especially if it were to mutate.	Perceived susceptibility	Perceived Severity Cues to Action	Likely to engage with prevention behaviours.
Fear of Travel	Travelling abroad triggered specific fears and people may have altered their plans. Perceived seriousness is high and perceived susceptibility may increase so may influence whether users would have travelled.	Perceived Severity Perceived Susceptibility	Perceived Severity Cues to Action	Less inclined to travel abroad.
Praying, Prayer, or call to God	Users prayed and called to God for family in West Africa. Users with family in West Africa may feel their family have high perceived risk.	Perceived Severity	Socioeconomic Knowledge Perceived Threat Cues to Action	More inclined to pray for people affected.
Symptoms	Users felt afraid if they were suffering from ill health symptoms. The perceived severity of Ebola is high so users may have thought they were suffering from Ebola when they experienced symptoms.	Perceived Severity	Socioeconomic Knowledge Perceived Threat Cues to Action	N/A
Quarantine	Users requested that those users who travel to West Africa and return should be quarantined. The perceived severity of Ebola is high so as an attempt to mitigate these threat users may have called for quarantine.	Perceived Severity	Socioeconomic Knowledge Perceived Threat Cues to Action	N/A

As can be seen from Table 7-7, the HBM was useful in predicting and considering how Twitter users were likely to behave based on information they were receiving from Twitter. However, one of the difficulties of such models is that human decision making and behaviour is complex and the HBM may not always be able to accurately predict how a person was likely to behave.

7.6 Utility of Information Theory

The concepts related to information theory could have been applied to the content of themes, because users were sharing and consuming information. However, there were certain themes where concepts related to information theory were particularly useful. In the themes 'fear of travel', Twitter users may have a number of information needs related to travel; moreover, these information needs tended to have been unmet when users were planning to travel. In the theme 'prevalence monitoring', information needs would have arisen because Twitter users wanted information on the precise location of a disease.

Certain users wanted to know how far an infectious disease was from their current location, and they would share this information on Twitter. With regard to prevention, a number information needs arouse as Twitter users were sharing tips on how to avoid a disease and this information may have been shared because it was highly sought after. For the themes related to media organisations, Twitter users were sharing and tweeting information from media organisations because these articles may have addressed the dormant information needs of Twitter users. For the theme Obama, Twitter users had information needs around the safety of the US president. There were also information needs that arose that were related specifically to the country of origin across the cases. There were also a number of information needs that arose surrounding the consumption of pork, and users would express that they had very specific information needs on whether particular food products were capable of transmitting a disease. There were also specific information needs that arose when Twitter users had questions surrounding medications and vaccines, as Twitter users required up to date information on the availability of specific medications.

7.7 Moral Panic Applied to Results

In discussing the results of this study at workshops and conferences, sociologists have noted that the response of Twitter users during both cases could be said to mirror a 'moral panic'. The concept of the moral panic was explained in section 2.6. A moral panic is when:

"A condition, episode, person or group of persons emerges to become defined as a threat to societal values and interests (Cohen, 2002: p.1)"

It could be argued that during the peak of the Ebola, swine flu, and Zika outbreaks a moral panic was underway, and Twitter users were caught up in this. The tweets from the fear theme can be used in support of this argument. In fact, there appeared to be an exaggerated fear from Twitter users across the themes, particularly in the discussions surrounding the possibility of patients rising from the dead, and of the potential of a zombie apocalypse. This reaction on Twitter coupled with increased media attention surrounding the outbreak could be argued to express a moral panic.

Another defining factor of moral panic is the exaggeration of an episode by mass media, and in the outbreaks of swine flu and Ebola, there were articles shared on Twitter that appeared to sensationalise the outbreaks. Looking back now at the outbreaks, sometime after they have occurred, it is easy to say that Twitter users, the general public, and the media might have over-reacted, but at those times some of the over-reaction might have been seen to be appropriate in order for people to become aware of a potential global threat. Nevertheless, it is worth mentioning this sociological concept in the context of the results of this study.

7.8 Outbreak Narrative Applied to Results

The concept of the outbreak narrative was explained in section 2.6, and this section will consider the applicability of the concept to the results of this study.

In the results of this present study, it was interesting to note the frequency of references to popular culture, zombies, and the potential of a zombie apocalypse. These surprising finding were unreported in previous evidence-based research on Twitter in the context of the outbreaks. It is worth drawing attention on media and culture literature in order to understand these phenomena.

In certain cases, it might be in the best interest of the media, especially unregulated online websites, to offer shocking narratives intentionally in order to generate page views and link clicks. This was found to be the case with the Ebola outbreak as it could be argued the news article on Ebola patients rising from the dead was intended to shock readers. Ostherr (2005) and Wald (2007) theoretically noted that outbreak narratives will be able to have an influence on the opinions of the public, and the present study provides empirical evidence that Hollywood outbreak narratives were indeed able to influence how some Twitter users were experiencing and understanding the outbreak. Ostherr (2005) and Wald (2007) noted that the narratives displayed in television and film might lead to negative public reactions to pandemics, and the media reports which are excessive and/or pessimistic might cause members of the public to become fearful. Importantly, they noted that narratives that provide users with information that is inaccurate might negatively affect the behaviour of the public and especially of individuals who may have limited or no formal education. Therefore, health authorities and governments should be aware of narratives surrounding infectious disease outbreaks, and respond swiftly and appropriately when a potential harmful narrative is shared. Twitter data could act as a platform that could be studied by those from disciplines such as English literature and/or history whom may not necessarily have research training in conducting surveys or interviews.

7.9 Utility of Twitter for Qualitative Research

This study utilised Twitter data to provide in-depth qualitative insights into the swine flu, Ebola and Zika outbreaks and is, therefore, in a unique position to comment on the effectiveness of utilising Twitter data for gaining qualitative insights. In the current body of literature, it appeared that the field of health evidence-based research has validated the use of Twitter for gaining real-time insights (Sinnenberg, 2017); however, surprisingly there is very little work examining the effectiveness of Twitter for providing in-depth qualitative insights into health topics. The difference between using quantitative methods, over qualitative methods, on Twitter data is that it may involve finding correlations between tweets and incidence rates, or that a large number of tweets are automatically processed and frequently occurring keywords are displayed. Qualitative Twitter studies are those that provide insight into the tweets within the context in which they have been written, by using research methods associated with qualitative research: only a limited number of studies have utilised in-depth qualitative methods to analyse Twitter data (Burch, Frederick, Pegoraro, 2015; Hewis 2015; Shepherd, Sanders, Doyle, and Shaw 2015).

The use of Twitter for gaining in-depth insights for qualitative research more generally was noted by Marwick (2013) in an article entitled *Ethnographic Qualitative Research on Twitter,* which summarised a number of textual analyses methods. However, Marwick (2013) offered no critical discussion of whether Twitter is an effective tool to conduct qualitative research. In 2017, Sinnenberg published a systematic review examining Twitter as a tool for health research. A total of 137 studies were considered, and it was found that 56% of them utilised content analysis to analyse data, 26% utilised surveillance methods, 14% engagement methods, 7% used Twitter for recruitment, 7% were based on intervention, and 4% were related to network analysis (Sinnenberg, 2017). The author noted that there should be standardised reporting guidelines and polices which reflect more on the ethical issues associated with researching Twitter. However, the study did not consider whether Twitter could be utilised to gain in-depth insights into health. Moreover, the study did not distinguish content analysis and thematic analysis when assessing the types of methods that were used to analyse Twitter data. This is important because thematic analysis is more associated with the qualitative approach.

A systematic review published in 2015 by Hu examined health communication research on digital platforms. This study found that research that utilises quantitative methods is more prevalent with 67% of published studies in comparison to 20.1% of articles which utilise qualitative research methods (Hu, 2015). Therefore, much of the previous research on Twitter may have been based on quantitative methods, which could explain why previous health research has not examined whether Twitter data has the potential to provide in-depth insights into health. The question of whether Twitter data could provide in-depth qualitative insights was considered in the field of marketing when Twitter was emerging as a platform for academic research (Branthwaite and Patterson, 2011).

Social media platforms provide those from the field of marketing with the ability to gather insights into how consumers might speak about products, and this is particularly useful when new products are launched. A number of platform-based, real-time analytic tools have been built which allow organisations to monitor discussions that take place on Twitter. Therefore, it is not surprising that one of the earliest and, to the best of the researcher's knowledge, only critical consideration of using social media data for qualitative research was published in 2011 in the Qualitative *Market Research Journal.* The aim of that article was to systematically study approaches to qualitative social media research. In the study, Branthwaite and Patterson (2011) noted that qualitative social media research is observational and therefore similar to ethnographic research, as well as quantitative surveys. The authors noted that the strength of utilising social media data is due to the immediacy of it, as it makes it possible to retrieve data in almost real-time. These statements are particularly true for the present study as it was observational and similar to the results of quantitative surveys. Moreover, it must be noted that, particularly from a public health perspective, health organisations would also benefit of the speed in which data can be collected and quantitatively analysed and which could provide important insights into the discussions that take place. In terms of limitations of using social media data for qualitative research, Branthwaite and Patterson (2011) noted that social media allow people to show an idealised version of themselves, and people might tend to exaggerate their views. Moreover, they noted that social media users might shift between two worlds: 1) offline everyday world, and 2) the social media world.

Overall, the study by Branthwaite and Patterson (2011) concluded that social media are interesting as they provide insights into social phenomena and running commentaries on everyday events. However, the authors noted that social media was not a suitable replacement for traditional in-depth qualitative research methods. The limitations of utilising Twitter data to gain in-depth insights in this present study were that 1) tweets are short and this could make it difficult to interpret the meaning of users; 2) there was a large amount of content that was not relevant and this might be avoided in traditional qualitative research, for instance in a survey or interview where there are opportunities to steer the questions on topic; 3) performing a textual analysis was a difficult and time-consuming task, and could be more difficult than coding an interview or a survey because of the volume of tweets and the number of non-relevant tweets; 4) in order to protect the anonymity of Twitter users, tweets had to be carefully anonymised, resulting in a time-consuming process that could also potentially alter the meaning of the tweets; 5) a lack of accurate demographic data, which might be available in traditional in-depth methods such as surveys or interviews. The table below was created using content from Branthwaite and Patterson (2011).

Table 7-8 Everyday world *vs*. social media world

Everyday world	Social media world
The everyday world is pragmatic, it has order and it is constrained.	The social media world might be self-indulging, and impractical.
The real world has as rules and conventions.	The goals people will have will be short-term and immediate.
The real world will be collaborative, cooperative which will be people to achieve personal, social, and emotional out.	The social media world might be superficial, momentary.
Goals will be long term, and they will evolve gradually based on consent and social influence.	Allows the creation of imaginary scenarios and fantasies.
EGO drive.	ID driven.

It is useful to reflect on the differences between the online and offline worlds, as shown in table 7-8 above and argued by Branthwaite and Patterson (2011). This means that tweets are posted in the online world governed and influenced by platform conventions. The argument here by Branthwaite and Patterson (2011) is that people who post on social media do so in an entirely different reality, which may make it difficult to treat social media data in the same light as survey or interview data. In this study, it was found, for example, that the language tended to be exaggerated and there was some evidence of situations and scenarios that tended to be imaginary. For these reasons, as well as the limitations outlined in the paragraph above, Twitter data might not be a good substitute to traditional qualitative research data, generated through interviews, etc. Indeed, it may also not be fair and/or even legitimate to compare traditional qualitative research methods, and data to those from social media. Social media platforms such as Twitter offer complimentary insights into human thoughts on a scale that was not possible before the advent of these platforms. A benefit of using Twitter for health research, as found in this study, is that a number of discussions emerged on Twitter for infectious diseases not reported in previous work. For instance, the rate at which popular culture was referred to. The difference between the online and offline world outlined by Branthwaite and Patterson (2011) may also be simplistic because researchers (Leander and McKim, 2003) now consider that the boundary between the online and offline world have now become increasingly blurred and that there are not two separate worlds but worlds that considerably overlap. Moreover, social media platforms should be studied by public health researchers using in-depth methods because of the influence social media can have over members of the public and due the vast number of users which are using these platforms . In-depth methods allow themes to emerge that may not be possible to extract or know outside of these platforms and which potentially would not come to light in a survey and/or interview.Therefore, although Twitter may not be a substitute for traditional qualitative research, it is an important source of data and should be studied.

7.10 Summary

The purpose of this chapter was to explore whether there were any potential similarities and differences between the results of the individual case studies from swine flu and Ebola as well as Zika. It was found that, although the Ebola and swine flu outbreaks had different symptoms, originated in different parts of the world and occurred five years apart, there were many similarities in how users responded. The comparative analysis, therefore, has highlighted potential similarities in the ways in which users were responding and commenting on events during the peak of the swine flu Pandemic from 2009, the Ebola epidemic from 2014 as well as the Zika outbreak from 2016. The study also found a number of similarities to emerge with previous empirical research, and also a number of themes that emerged in this present study that had not been reported in previous empirical work. The utility and applicability of concepts surrounding Information Theory, and the Health Belief Model were also considered.

Chapter 8 – Conclusion

8.1 Introduction

The previous chapters of this study described and discussed the topic of this thesis, which sought to develop a better understanding into how Twitter users communicate about infectious disease outbreaks by providing an in-depth analysis of tweets related to the 2009 swine flu, the 2014 Ebola outbreak, and the 2016 Zika outbreak. This chapter provides a conclusion to the overall study by demonstrating how the research questions were addressed (section 8.2), the implications of the results (section 8.3), the limitations of this research (section 8.4), the strengths (section 8.5), the contributions to knowledge (section 8.6). It provides suggestions for future research (section 8.7), and finally, it provides a summary of the thesis (section 8.8).

8.2 Research Questions

In Introduction Chapter One (section 1.4) a number of research questions were proposed, and these are summarised below:

Q1. What does the analysis of social media data such as Twitter tell us about how people communicate about disease outbreaks on Twitter?

- It was found that the discussion on Twitter during infectious disease outbreaks varies around a number of different topics such as current events, users' thoughts and feelings.

  Some of these discussions were based on factual information; however, there were also a number of discussions taking place that were based on fantasies, such as when Twitter users referred to popular culture in their tweets.
- More specifically, discussion around the swine flu outbreak in 2009 during the peak was based on seven key themes as described below:
  - The theme 'emotion and feeling' consisted of users who expressed emotion towards the swine flu outbreak and included sub-themes of fear, worry, anger, and panic. The theme 'health related' included tweets which discussed medical concepts related to transmission, prevention techniques, symptoms, medication, and diagnosis. There was a theme entitled 'general commentary and resources', which expressed general comments concerning the swine flu outbreak. A theme 'media and health organisations' also emerged, which

included tweets that mentioned a media organisation and/or expressed a view towards the media. A theme on politics was also emerged and this was where Twitter users made references to politics and/or a political figure. There was also a theme entitled country of origin where users referred to Mexico, Mexicans, and travel. A further theme to emerge was food, which referred to bacon, or which referenced food in other ways by mentioning kosher, and halal meat and this theme included users who mentioned vegan and vegetarian diets. Another interesting theme to emerge was related to humor and sarcasm, where Twitter users shared tweets that expressed humor and sarcasm towards swine flu.

The discussion surrounding the Ebola outbreak during its peak of 2014 was based on eight key themes as described below:

- An emergent theme was that in which users expressed emotion towards Ebola and included themes of anger, fear, fear of travel, praying, prayers, and calls to God. A further theme to emerge was 'health-related information', where Twitter users discussed medical concepts such as transmission, prevention, symptoms and vaccines. Another theme to emerge was Twitter users who would tweet about news stories which appeared to be significant during this time. A further theme that emerged mentioned official organisations, health organisations, and health charities in general terms. Another theme that emerged was one in which Twitter users referred to regions that were specific to West Africa, such as Sierra Leone. Another interesting theme to emerge was centered around politics where Twitter users referred to politics and/or a political figure and where Twitter users were critical of the government. There was also a theme, which was based on humour and/or sarcasm.

The discussion surrounding the Zika outbreak during its peak of 2016 was based on key themes as described below:

- Twitter users would discuss concepts surrounding pregnancy and share links to news articles and debates around abortion would arise. There was also a travel and Olympics theme which contained tweets from users who were sharing articles mentioning the Olympic games taking place at the time. A further theme that emerged would contain references to conspiracy theories such as that genetically modified mosquitoes were responsible for the re-emergence of Zika. There were also tweets mentioning and criticizing the WHO for their role in the outbreak. A number of users would also share general health information such as the symptoms and prevention methods associated with Zika. A further theme would emerge related to tracking the Zika virus. There was also a more general theme which contained tweets which would discuss the Zika outbreak in broad terms.

Importantly, among the themes identified, this present study uncovered a substantial number of discussions taking place not reported in previous research as highlighted in section 6.3.9, section 6.3.10, and section 7.3.1.

Q2. What similarities and differences emerge when comparing how people respond to infectious disease outbreaks on Twitter?

- As highlighted in Discussion Chapter 7 section 7.3.8 there were a number of similarities and differences between the cases. The similarities were based on the following themes: fear, anger, fear of travel, transmission, prevalence monitoring, speculative diagnosis, prevention, symptoms, medications (e.g. vaccines), economic impact of disease, information seeking, voice of reason or downplaying the outbreak, general discussions, references to official organisations, references to Obama, references to areas affected, humour, sarcasm, and references to popular culture. This comparison between the two cases uncovered reactions that could be found to emerge in future infectious disease outbreaks.
- There were, however, themes that emerged only for swine flu and there were themes that were specific to Ebola. The most striking differences between the two cases were that, for swine flu, Twitter users would tweet about whether it was safe to consume pork products and for Ebola there were tweets around a news story which suggested Ebola patients had risen from the dead, and there were discussions which centered around sharing information related to conspiracy theories.
- There were a number of similarities to the response of users related to Zika such as fear, mentions of the WHO, political references, transmission, prevention, and travel. Twitter users also discussed the name of Zika, as was the case with swine flu. However, for the Zika epidemic there were a number of discussions that were specific to Zika such as: avoid pregnancy narrative, Zika threat to pregnant women, Zika virus spreads fear among pregnant Brazilians, Zika threat to pregnant Columbians, abortion debate, pregnancy, Olympics Rio 2016, Mosquitoes, Mosquitoes, Microcephaly, Zika Spreading

explosively in South and Central America, Zika will spread across the Americas, and a theme which mentioned Brazil and the Zika Virus.

Q3. What is the potential of Twitter to provide in-depth insights into how citizens communicate about infectious disease outbreaks?

- Social media research is a rapidly developing and emerging field, and as was noted in Chapter 2 Literature Review section 2.14 previous studies had utilised Twitter for health purposes were primarily making use of quantitative techniques. One of the aims of this present study was, therefore, to develop a critical understanding of whether Twitter could provide in depth insights into health topics.
- As discussed in section 6.9, a key strength of Twitter data is that it is available very quickly; if captured in real time there is little cost, and it can be utilised to gain a better understanding of the types of discussions that are taking place during an unfolding infectious disease outbreak.
- Overall, however, it was found that tweets lack the depth that might be present in a traditional qualitative study (e.g. interviews, focus groups) and no demographic data are available for tweets which might be available in traditional qualitative methods such as surveys and interviews. Moreover, there are important differences in how people behave online, as opposed to offline, which might mean that the data are not necessarily comparable to those generated in traditional qualitative research. This is a substantial contribution to knowledge because in the field of health previous empirical based research has approached this question.
- Although it was found that Twitter data does offer insights into human behaviors and thoughts at very little cost and should be studied as a complimentary source of research data.

Q.4 What characteristics have enabled Twitter to thrive for health-related research in comparison to other social media platforms?

- As outlined in section 2.7.1, Twitter data can be retrieved at little to no cost if it is captured in real-time and Twitter is the only platform to make its data available on this scale. Therefore, a potential reason for its popularity among health research could relate to the availability of data.
- As found by reviewing literature in this area, an interesting aspect of Twitter that could be mined for data analysis is that the platform also can be used for public health

campaigns and initiatives. This could also form a further reason for the popularity of the platform.

- Tweets also tend to be short which makes it easier for researchers to analyses the data in comparison to platforms such as Facebook which have a higher character count.

8.3 Implications of Results

The study made use of thematic analysis, which involves reading and labelling individual tweets, and it was found that, although time-consuming, this method allowed many new themes and discussions to emerge that were taking place on Twitter that had not been reported in previous literature.

An implication of this present study, therefore, is that for researchers and public health officials wishing to gain greater insights, a potential method of analysis could be to utilise in-depth qualitative methodologies. Several of the results that were reported could be used by health authorities when disseminating information via social media platforms. For example, the results reported surrounding the fear of travel could potentially be used when disseminating information on social media platforms, such as Twitter, because health authorities could provide information to users on any potential risks that they may face by travelling to, or in, affected areas. In future infectious disease outbreaks, it may be possible to monitor Twitter for tweets that potentially express anger, and to disseminate tailored information, for example, stress reduction strategies and suggestions for users who may be feeling uneasy due to the outbreak. The results around transmission and prevalence monitoring could potentially be used by health authorities when disseminating information on Twitter, for example, by providing real-time updates and online tracking capabilities. Additionally, health authorities may wish to highlight how users should use this information and whether users should be alarmed if a disease is reported close to their own geographical location. Health authorities may be interested in the results in relation to the prevention of swine flu, as they may have been disseminating prevention information on the swine flu outbreak on the platform during the outbreak. The information shared by users could be cross-referenced with official guidance in order to ascertain whether the information shared by Twitter users is factual.

It also appears that the name ‘swine flu’ given to the A/H1N1pdm09 virus caused confusion among Twitter users. This information could be used by governments, health authorities, and food safety bodies to pressure media outlets to refer to swine flu by its medical term. In future disease outbreaks, in case of confusion over the name of a disease, health authorities should rapidly disseminate information that corrects any potential misinformation about this issue.

Additionally, the findings surrounding the rate at which Twitter users were seeking information may be of interest to health authorities. This is because in future outbreaks of infectious diseases, health authorities may wish to monitor discussions on Twitter and respond to users who may be seeking information. For example, if a Twitter user sought information on how to differentiate between swine flu and the regular flu, then a health authority could potentially respond to the tweet in order to address a user's information needs.

In these case studies, it appeared that narratives surrounding infectious disease outbreaks were able to influence how certain Twitter users were perceiving them. This led to references to popular culture, such as zombies and zombie apocalypse. A specific news story shared and discussed during the Ebola outbreak indicated that a patient had risen from the dead. These narratives could potentially be harmful and health authorities, governments and policy makers should be aware of this and provide advice and guidance on the accuracy of such narratives.

8.4 Implications for Social Media Research

The results of this thesis have implications towards wider research on social media across a number of different disciplines. This is because this study has shown the potential of in-depth qualitative methodologies to extract greater insight into data derived from social media platforms in comparison to quantitative methods. These findings could potentially help in informing researchers seeking to analyse data from other social media platforms such as Facebook because by employing an in-depth qualitative method may allow for greater depth in the results. In addition to this, a wider implication is the potential of social media to uncover public views and opinions on health that may not necessarily emerge in an interview and/or survey, for example, this could occur due to the potential of interview bias. Moreover, the results have also uncovered important insights into how users converse about health across Twitter and how users may draw on other social media platforms during outbreak situations and link to these on other platforms.

However, one of the difficulties of researching other social media platforms such as Facebook, Instagram, and/or LinkedIn is that data from these platforms is not available to researchers as is the case with Twitter data. On Facebook, for instance, much of the discussion that takes place occurs privately between friends and the majority of accounts on Facebook are only visible to the immediate friends of a user. Facebook may also contain groups where discussion may take place, however, these groups are often closed and there is no Application Programming Interface (API) for researchers to retrieve data. One of the strengths, therefore, of researching Twitter is that almost all of the data generated on the platform can be retrieved by researchers. Therefore, it may not be possible to conduct a study of this scale on other social media platforms due to the difficulties of accessing data. There may also be intrinsic differences on how consumers engage with different social media platforms and Alhabash & Ma (2017), for instance, found that there were differences in the use of Facebook, Twitter, Instagram, and Snapchat among college students.

8.5 Limitations of the Research

The study examined two-day time intervals from when there was heightened interest surrounding swine flu, Ebola, and Zika; therefore, there could be limitations in the conclusions that were drawn from the data. However, evidence was found in each of the chapters for the newly identified themes to have occurred throughout the respective outbreaks.

This study concentrated on examining tweets in English, and therefore it is not a complete record of all users that were tweeting about the outbreaks as other languages were not considered. Thus, a limitation of the study is that it may not have brought in the voices of Twitter users who may have been on the front-line during the outbreaks of swine flu, Ebola, and Zika studied in this thesis and who may have been tweeting in French, Spanish and/or Portuguese, for instance. However, tweets in English dominate Twitter and users from other countries are likely to retweet and engage with non-English tweets. For instance, in the Ebola study 80% and upwards of tweets were sent and received in English. Moreover, the focus of this research was to develop a better understanding of the content shared on the platform during the peak of the respective outbreaks. Furthermore, as the majority of content on Twitter is based on tweets in the English-language a broader international perspective is likely to still have value. This is because the results of the empirical case studies on swine flu, Ebola, and Zika identified important trends in user responses to infectious disease outbreaks.

A limitation related to retrieving data via keywords is that the study may not have retrieved all data from Twitter related to the outbreaks because certain users may have been talking about the outbreaks without mentioning them, although this is perhaps unlikely to occur frequently. Moreover, it was also found that certain users may have mentioned the keywords used to retrieve data without commenting on the current outbreaks and this was highlighted in section 3.5.1 Twitter as a source of data. Tweets may have been sent on either of the outbreaks which were retrieved in this study but which contained irrelevant content, and also tweets may have been received which would contain only a minor relationship to the topic that was analysed. However, this is a wider limitation of conducting research on social because unlike in a survey and/or interview it is not possible to guide users to respond in certain ways. Likewise, this is also a strength because it meant that Twitter users would discuss aspects of the outbreaks they may not feel comfortable in disclosing during a formal academic survey and/or interview.

When extracts from each of the tweets were provided, they were carefully reworded, therefore, the original intended meaning from users may have been lost; however, if was deemed more important to protect the identities of Twitter users from an ethical point of view. It must also be noted that data was captured and analysed on the outbreaks after considerable time had elapsed which is a potential limitation of this study. The study analysed tweet content and did not categorize and/or analyse images or videos in tweets.

In regards to the findings It may appear that certain themes have low number of tweets assigned to them and, for readers not familiar with social media research, it may appear that some of the percentages associated with each of the themes are low. However, this is a common feature of research on social media because there will be substantial non-relevant content shared by automated accounts (Chu, Gianvecchio, Wang, & Jajodia 2010; Wald, Khoshgoftaar, Napolitano, and Sumner, 2013; Ferrara, Varol, Davis, Menczer, and Flammini, 2016). Previous qualitative Twitter studies also report percentages and numbers of similar magnitude (Chew and Eysenbach, 2010; Burch, Frederick & Pegoraro, 2015). Moreover, although a theme appears to have a low amount of tweets, the tweets included in it could still have been popular on Twitter because they might have been viewed widely (tweet activity dashboard, n.d.).

8.6 Strengths of the Research

This study is the first empirical study to utilise Twitter data from the peak of infectious disease outbreaks, to compare these to one another to ascertain commonalities and differences, and then to consider whether Twitter can be utilised to provide qualitative insights for public health research. Moreover, to the best of the author's knowledge, it is the largest qualitative study undertaken on Twitter data related to infectious disease outbreaks. A further strength of the study is that it provided three detailed qualitative reports (Chapter 4 Swine Flu section 4.7, Chapter 5 Ebola section 5.6, and Chapter 6 Zika, section 5.5), which included tweet illustrations, into how Twitter users responded during the peak of the swine flu, and Ebola outbreaks respectively. This study provided descriptions of the Health Belief Model and concepts surrounding Information Theory, which were then applied in the interpretation of results providing the study with greater interpretation potential. Moreover, the study also drew on sociological concepts such as the Outbreak Narrative and the Moral Panic which were also applied to the results and this is the first empirical study to link these concepts to user responses on Twitter. Previous studies on Twitter have used the free API and then taken a sample of data (i.e. a sample of a sample). A strength of this study is that for the study on swine flu and Ebola it retrieved firehose data, systematically eliminated popular content, and only then took a sample. These steps were carefully reported in the methodology section 3.13.

8.7 Contribution to new knowledge

This study provided new insights into how users responded during peaks of the 2009, swine flu pandemic, the 2014 Ebola epidemic, and the 2016 Zika epidemic. Moreover, the study overall contributes the current body of knowledge by reflecting on the potential of Twitter data for providing in-depth qualitative insights. Of equal importance, it also reported on a number of similarities and differences on how Twitter users responded during the respective outbreaks. The literature review provided a theoretical contribution as a number of studies that utilised Twitter data were brought together and compared (Chapter 2, Literature Review, Sections 2.11 to 2.14). A substantial contribution to knowledge is based on the results of this study, i.e. the in-depth thematic analysis of tweets (Chapter 4 Swine Flu section 4.7, Chapter 5 Ebola section 5.6, Chapter 6 Zika, section 5.5), where a number of new results were reported which are likely to be of interest organisation such as the World Health Organisation (WHO). Another theoretical contribution in these chapters related to the applicability of Information theory and the Health Belief Model to the results of the study, i.e. Twitter data (Discussion Chapter 7 section 7.5 and section 7.6). The study also contributed to knowledge by linking the sociological concepts of the Outbreak Narrative, and the Moral Panic to infectious disease outbreaks on Twitter (Discussion Chapter 7 section 8.7 and section 8.8). The methodological contribution of the study is related to the steps that were taken in order to source and filter Twitter data in a way which made performing an in-depth analysis feasible (Methodology

Chapter 3, section 3.14). Another theoretical and methodological contribution is based on the reflection of whether Twitter can be utilised in place of traditional methods for gaining in-depth insights (Discussion Chapter 7 section 7.9).

8.8 Suggestions for Further Research

A similar study could be conducted utilising meta-data such as location information in order to associate themes with the locations they were being tweeted from. This would allow researchers to explore whether specific locations influenced tweets, i.e. whether those that were most affected had different sentiments and/or discussions. Moreover, this would allow for the ability to compare broader information from Twitter to those users who were on the front-line and whom may have been tweeting in different languages. Further research could sample Twitter data over a longer period of time and/or examine multiple days across the outbreak scenarios in order to see whether specific themes arose on certain days or whether they were occurring throughout the whole outbreak. Data from a number of different languages related to infectious disease outbreaks could be retrieved and included in the analysis of tweets. Further research could also attempt to derive important themes, or those which were shared by influential Twitter users by using the influence scores of users, and network analysis. In addition to performing a textual analysis of tweets, further research could also examine Web links that were shared as well as media, such as images and videos. It would be also interesting to compare the results from Twitter to those that were conducted with the offline population. A more comprehensive study could seek to work with Twitter in order to retrieve all tweets that were posted over the days from when there was heightened interest surrounding the outbreaks. Further research could try to understand whether the information needs of those users asking questions during outbreaks were addressed and, if so, whether the information they were provided was accurate and trustworthy.

8.9 Thesis Summary

This chapter has provided a conclusion to the overall study by summarising the research questions, contributions to knowledge, strengths, limitations, and finally by providing a number of potential recommendations for future research. Reyneke, Pitt, and Berthon (2012) have noted that social media platforms can be considered to be just as influential as conventional media. Infectious disease outbreaks are a serious public health concern, and platforms such as Twitter are one of the mechanisms with communicating with the public when they occur. It was, therefore, important to study the types of content that was shared on the platform during three recent infectious disease outbreaks of swine flu, Ebola, and Zika. During infectious disease outbreaks, specific information needs may arise. Public health authorities should be aware of this and disseminate appropriate information. In Chapter 7 Discussion, section 7.4 provided guidance on potential information which could be disseminated during an infectious disease outbreak. In addition to generic information needs, there may also be specific information needs that arise related to an outbreak (e.g. related to the geographical location of origin); as such, it is important for health authorities to be aware of this and discriminate when addressing the general public. Overall, the thesis has conducted an in-depth thematic analysis of tweets related to three different infectious disease outbreaks in order to extract individual insights and to examine potential commonalities and differences in themes.

References

About Twitter. (2016). Twitter Usage/Company Facts. Retrieved 19 June, 2017, from https://about.twitter.com/company.

Aest, A. E. (2009, April 29). Egypt orders pig cull. *ABC News.* Retrieved from http://www.abc.net.au/news/2009-04-30/egypt-orders-pig-cull/1667152 [Last Accessed 09/03/17].

Ahmed, W., & Bath, P. A. (2015). *Comparison of Twitter APIs and tools for analysing Tweets related to the Ebola Virus Disease.* Presented at iFutures. Sheffield, 07 July 2015.

Ahmed, W., Demartini, G., & Bath, P. A. (2017). *Topics Discussed on Twitter at the Beginning of the 2014 Ebola Epidemic in United States.* Presented at iConference 2017. Wuhan, China, March 2017.

Aiken, A. R., Scott, J. G., Gomperts, R., Trussell, J., Worrell, M., & Aiken, C. E. (2016). Requests for abortion in Latin America related to concern about Zika virus exposure. *New England Journal of Medicine*, *375*(4), 396-398.

Ali, N. S. (2002). Prediction of coronary heart disease preventive behaviours in women: A test of the Health Belief Model. *Women & Health, 35*(1), 83–96.

Alhabash, S., & Ma, M. (2017). A Tale of Four Platforms: Motivations and Uses of Facebook, Twitter, Instagram, and Snapchat Among College Students?. *Social Media+ Society*, *3*(1), 2056305117691544.

American Psychiatric Association. (2013). *American Psychiatric Association (APA), Diagnostic and statistical manual of mental disorders* (5th ed.)*.* Washington, DC: APA.

Arabghalizi, T., & Rahdari, B. (2017). Event-based user profiling in social media using data mining approaches. *POLITesi.*

Arthur, C. (2014, March 27). Google Flu Trends is no longer good at predicting flu, scientists find. *The Guardian*. Retrieved from http://www.theguardian.com/technology/2014/mar/27/google-flu-trends-predicting-flu.

Baize, S., Pannetier, D., Oestereich, L., Rieger, T., Koivogui, L., Magassouba, N., & Günther, S. (2014). Emergence of Zaire Ebola Virus Disease in Guinea – Preliminary Report. *The New England Journal of Medicine*, 1–8. doi: 10.1056/NEJMoa1404505.

Bandura, A. (1986). Social Foundations of Thought and Action: A Social Cognitive Theory. *Englewood Cliffs, NJ: Prentice-Hall.*

Bandura, A. (1977). Self-efficacy: Toward a unifying theory of behavioural change. *Psychological Review, 84*, 191–215.

Bandura, A. (1997). *Self-efficacy: The exercise of control*. London: Macmillan.

Brasil, P., Pereira Jr, J. P., Moreira, M. E., Ribeiro Nogueira, R. M., Damasceno, L., Wakimoto, M., ... & Zin, A. A. (2016). Zika virus infection in pregnant women in Rio de Janeiro. *N Engl J Med*, *2016*(375), 2321-2334.

BBC News (2016, January 14). Ebola: Mapping the outbreak. Retreived from http://www.bbc.co.uk/news/world-africa-28755033.

Becker, M. H. (ed). (1974). The health belief model and personal health behaviour. *Health Education Monographs, 2*, 324–508.

Belkin, N. J., & Vickery, A. (1989). *Interaction in Information Systems: A Review of Research from Document Retrieval to Knowledge-Based Systems*. London: British Library.

Bengtsson, M. (2016). How to plan and perform a qualitative study using content analysis. *NursingPlus Open, 2*, 8–14.

Beninger, K., Fry, A., Jago, N., Lepps, H., Nass, L., & Silvester, H. (2014). *Research using social media; users' views.* London: NatCen Social Research. Retrieved from http://www.natcen.ac.uk/media/282288/p0639-research-using-social-media-report-final-190214.pdf.

Bernardo, T. M., Rajic, A., Young, I., Robiadek, K., Pham, M. T., & Funk, J. A. (2013). Scoping review on search queries and social media for disease surveillance: A chronology of innovation. *Journal of Medical Internet Research, 15*. doi: 10.2196/jmir.2740.

Beutler, B. (2009, April 30). Top Conservatives On Twitter FAIL. *TMP DC.* Retrieved from http://talkingpointsmemo.com/dc/top-conservatives-on-twitter-fail.

Bhattacharya, S., Srinivasan, P., & Polgreen, P. (2014). Engagement with health agencies on Twitter. *PloS ONE, 9*(11). doi: 10.1371/journal.pone.0112235.

Boeije, H. (2009). *Analysis in qualitative research*. California: Sage publications.

Borra, E., & Rieder, B., (2014). Programmed method: Developing a toolset for capturing and analyzing tweets. *Aslib Journal of Information Management, 66* (3), pp.262–278.

Bosley, J., Zhao, N., Hill, S., & Shofer, F. (2013). Decoding twitter: Surveillance and trends for cardiac arrest and resuscitation communication. Resuscitation. [Online] Retrieved from http://www.sciencedirect.com/science/article/pii/S0300957212008714 [Last Accessed 14/12/14].

Bostrom, N. (2008). *Global Catastrophic Risks*. Oxford: Oxford University Press.

Boyd, D., Scott, G., & Gilad, L. (2010*). Tweet, Tweet, Retweet: Conversational Aspects of Retweeting on Twitter*. *Proceedings HICSS-43, January 6*. Kauai: IEEE.

Boynton, G. R. (2014). Co-Motion—Twitter Enters the Public Domain. Open Journal of Political Science, *04*(02), 76–88. doi: 10.4236/ojps.2014.42009.

Branthwaite, A., & Patterson, S. (2011). The power of qualitative research in the era of social media. *Qualitative Market Research: An International Journal, 14*(4), 430–440.

Bredderman, W. (2016, August 4). 'No Doubt' President Obama Is a Muslim, Donald Trump's New York Campaign Chair Says. *Observer*. Retrieved from http://observer.com/2016/08/no-doubt-president-obama-is-a-muslim-donald-trumps-new-york-campaign-chair-says/.

Briand, S., Bertherat, E., Cox, P., Formenty, P., Kieny, M.-P., Myhre, J. K., & Dye, C. (2014). The International Ebola Emergency. *The New England Journal of Medicine*, 1–4. doi: 10.1056/NEJMp1409858.

Bright Planet. (2013, June 25). Twitter Firehose *vs*. Twitter API: What's the difference and why should you care? *Bright Planet.* Retrieved 12 March, 2015, from http://www.brightplanet.com/2013/06/twitter-firehose-vs-twitter-api-whats-the-difference-and-why-should-you-care/.

Brooker, P., Barnett, J., & Cribbin, T. (2016). Doing social media analytics. *Big Data & Society*, *3*(2).

Bruns, A., & Highfield, T. (2012a). *Blogs, Twitter, and breaking news: The produsage of citizen journalism. In Lind, Rebecca Ann (Ed.) Producing Theory in a Digital World: The Intersection of Audiences and Production in Contemporary Theory*. (pp. 15–32). New York: Peter Lang Publishing Inc.

Bruns, A., & Liang, Y. E. (2012b). Tools and methods for capturing Twitter data during natural disasters. *First Monday, 17*(4). ISSN: 1396-0466. Retrieved from http://firstmonday.org/ojs/index.php/fm/article/view/3937/3193.

Bruns, A., & Stieglitz, S. (2013). Towards more systematic Twitter analysis: Metrics for tweeting activities. *International Journal of Social Research Methodology, 16*, 91–108. doi: 10.1080/13645579.2012.756095.

Bryman, A. (2008). Social Research Methods. *Social Research*, *3*. doi: 10.4135/9781849209939.

Burch, L. M., Frederick, E. L., & Pegoraro, A. (2015) Kissing in the Carnage: An Examination of Framing on Twitter During the Vancouver Riots. *Journal of Broadcasting & Electronic Media, 59*(3), 399–415. doi: 10.1080/08838151.2015.1054999.

Burnap, P., Rana, O., Williams, M., Housley, W., Edwards, A., Morgan, J., ... Conejero, J. (2014) COSMOS: Towards an Integrated and Scalable Service for Analyzing Social Media on Demand. *International Journal of Parallel, Emergent and Distributed Systems. 30*(2), 80–100.

Cameron, M. A., Power, R., Robinson, B., & Yin, J. (2012). *Emergency situation awareness from twitter for crisis management. Proceedings of the 21st international conference companion on World Wide Web – WWW '12 Companion* (pp. 695). New York: ACM Press. doi: 10.1145/2187980.2188183.

Capurro, R., & Hjørland, B. (2005). The concept of information. *Annual Review of Information Science and Technology, 37*(1), 343–411. doi: 10.1002/aris.1440370109.

Carneiro, H. A., & Mylonakis, E. (2009). Google trends: A web-based tool for real-time surveillance of disease outbreaks*. Clinical infectious diseases, 49*(10), 1557–1564.

Case, D. O. (2012). *Looking for Information: A Survey of Research on Information Seeking, Needs and Behaviour, Part 1* (3rd ed.). West Yorkshire: Emerald Group Publishing.

Case, D. O. (2016). *Looking for Information: A Survey of Research on Information Seeking, Needs and Behaviour, Part 1* (4th ed.). West Yorkshire: Emerald Group Publishing.

Cavazos-Rehg, P. A., Krauss, M., Fisher, S. L., Salyer, P., Grucza, R. A., & Bierut, L. J. (2015). Twitter chatter about marijuana. *The Journal of Adolescent Health: Official Publication of the Society for Adolescent Medicine, 56*(2), 139–45. doi: 10.1016/j.jadohealth.2014.10.270.

Centres for Disease Control and Prevention [CDC]. (2010). Updated CDC estimates of 2009 H1N1 influenza cases, hospitalizations and deaths in the United States, April 2009 – April 10, 2010. Available from http://www.cdc.gov/h1n1flu/estimates_2009_h1n1.htm.

Centres for Disease Control and Prevention [CDC]. (2004). Program Operations Guidelines for STD Prevention: Community and Individual Behaviour Change Interventions. Retrieved 20 August, 2015, from http://www.cdc.gov/std/program/community/9-PGcommunity.htm.

Centres for Disease Control and Prevention [CDC]. (2015). Ebola (Ebola virus disease). Retrieved 22 January, 2015, from http://www.cdc.gov/vhf/ebola/transmission/.

Centres for Disease Control and Prevention (CDC). (2016). *Zika Travel Information*. Zika Travel Information. Retrieved 05 December, 2017, from https://wwwnc.cdc.gov/travel/page/zika-information.

Chaffee, S. H. (1991). *Communication Concepts, Explication.* Newbury Park: Sage publications.

Chapple, A., Ziebland, S., (2004). The role of humor for men with testicular cancer. *Qual. Health Res., 14*(8), 1123–1139. doi: 10.1177/1049732304267455.

Chew, C., & Eysenbach, G. (2010). Pandemics in the age of Twitter: Content analysis of tweets during the 2009 H1N1 outbreak. *PLoS ONE, 5*(11).

Cho, Y. (2008). Intercoder Reliability. In P. J. Lavrakas (Ed.), *Encyclopedia of Survey Research Methods. (*pp. 345–346). Thousand Oaks, CA: Sage Publications, Inc. doi: 10.4135/9781412963947.n228.

Chou, W. Y. S., Hunt, Y. M., Beckjord, E. B., Moser, R. P., & Hesse, B. W. (2009). Social media use in the United States: implications for health communication. *Journal of medical Internet research*, *11*(4).

Chorus. Project site for the Chorus Twitter analytics tool suite. Retrieved 23 December, 2014, from http://chorusanalytics.co.uk/.

Chu, Z., Gianvecchio, S., Wang, H., & Jajodia, S. (2010, December). *Who is tweeting on Twitter: Human, bot, or cyborg? Proceedings of the 26th annual computer security applications conference* (pp. 21–30). New York: ACM Press.

Coe, A. B., Gatewood, S. B. S., Moczygemba, L. R., Goode, J.-V. K. R., & Beckner, J. O. (2012). The use of the health belief model to assess predictors of intent to receive the novel (2009) H1N1 influenza vaccine. *Innovations in Pharmacy. 3*(2), 1–11.

Cohen, S. (2002). *Folk Devils and Moral Panics* (3rd edn). London: Routledge.

Constine, J. (2014, November 12). Twitter Will Onboard Users With "Instant Timeline", Inject Top Tweets From "While You Were Away". TechCrunch. Retrieved from https://techcrunch.com/2014/11/12/twitter-instant-timeline/ [Last Accessed 29/04/17]

Çoruh, B., Ayele, H., Pugh, M., & Mulligan, T. (2005). Does Religious Activity Improve Health Outcomes? A Critical Review of the Recent Literature. *EXPLORE: The Journal of Science and Healing. 1*(3), 186–191, doi: 10.1016/j.explore.2005.02.001.

Creswell, J. W. (2003). *Research design: Qualitative, quantitative, and mixed methods approaches* (2nd ed.) Thousand Oaks: Sage.

Davis, C. P. (2015). Swine Flu: Facts, Symptoms, Treatment And H1N1 Vaccine. *MedicineNet.* Retrieved 11 October, 2015, from http://www.bibme.org/citation-guide/apa/website.

De Quincey, E., & Kostkova, P. (2010). Early warning and outbreak detection using social networking websites: The potential of Twitter. *In Lecture Notes of the Institute for Computer Sciences, Social-Informatics and Telecommunications Engineering.* 27, 21–24. doi: 10.1007/978-3-642-11745-9_4.

Diaz, J. A., Griffith, R. A., Ng, J. J., Reinert, S. E., Friedmann, P. D., & Moulton, A. W. (2002). Patients' use of the Internet for medical information. *Journal of General Internal Medicine, 17*(3), 180–185.

Dillenburger, K., Jordan, J. A., McKerr, L., Devine, P., & Keenan, M. (2013). Awareness and knowledge of autism and autism interventions: A general population survey. *Research in Autism Spectrum Disorders, 7*(12), 1558–1567. doi: 10.1016/j.rasd.2013.09.004.

Discovertext.com (n.d.). Retrieved 12 February, 2015, from: https://www.discovertext.com/.

Social Media Ebola Hoax Causes Deaths. (2014) Retrieved 1 May, 2015, from http://news.discovery.com/human/psychology/social-media-ebola-hoax-causes-deaths-14100.htm.

Dredge, S. (2014, October 22). Twitter changes: 20 hits and misses from the social network's history. *The Guardian*. Retrieved from https://www.theguardian.com/technology/2014/oct/22/twitter-changes-hits-misses-history#img-2.

Duggan, M., Ellison, N.B., Lampe, C., Lenhart, A, and Madden, M. (2014). Social Media Update 2014. *Pew Research Center*. Retrieved 11 October, 2015, from http://www.pewinternet.org/2015/01/09/social-media-update-2014/.

Dumbrell, D., & Steele, R. (2013). Twitter and health in the Australian context: What types of information are health-related organizations tweeting? *Proceedings of the Annual Hawaii International Conference on System Sciences (HICSS).* Hawaii: IEEE.

Ekman, P. (1992). An argument for basic emotions. *Cognition and emotion, 6*(3–4), 169–200.

Elachola, H., Gozzer, E., Zhuo, J., & Memish, Z. A. (2016). A crucial time for public health preparedness: Zika virus and the 2016 Olympics, Umrah, and Hajj. *The Lancet*, *387*(10019), 630-632.

Eleftheriou-Smith, L.-M. (2014, October 6). Ebola 'risen from the dead' zombie story is a complete hoax'. *The Independent*. Retrieved from http://www.independent.co.uk/news/weird-news/ebola-risen-from-the-dead-viral-zombie-story-is-a-complete-hoax-9777916.html.

Eisenhardt, K. M. (1989). Building theories from case study research. *Academy of management review*, 14(4), 532-550.

Ellison, N. B. & Boyd, D. (2013). Sociality through Social Network Sites. In W. H. Dutton (Ed.), *The Oxford Handbook of Internet Studies* (pp. 151–172). Oxford: Oxford University Press.

Eversley, M. (2014, October 12). Trump hits Twitter to blame Obama for Ebola in NYC. *USA Today.* Retrieved from https://www.usatoday.com/story/news/nation/2014/10/23/trump-ebola-twitter-obama/17815841/.

Eysenbach, G., Diepgen, T. L., Gray, J. A. M., Bonati, M., Impicciatore, P., Pandolfini, C., & Arunachalam, S. (1998). Towards quality management of medical information on the internet: Evaluation, labelling, and filtering of information. *BMJ, 317*(7171), 1496–1502. doi: 10.1136/bmj.317.7171.1496.

Jin, F., Wang, W., Zhao, L., Dougherty, E., Cao, Y., Lu, C.-T., & Ramakrishnan, N. (2014). Misinformation Propagation in the Age of Twitter. *Computer. 47*(12) 90–94.

Fauci, A. S., & Morens, D. M. (2016). Zika virus in the Americas—yet another arbovirus threat. *New England Journal of Medicine*, *374*(7), 601-604.

Ferrara, E., Varol, O., Davis, C., Menczer, F., & Flammini, A. (2016). The rise of social bots. *Communications of the ACM, 59*(7), 96–104.

Fishbein, M. & Ajzen, I. (1975). *Belief attitude, intention and behaviour. An introduction to theory and research*. Menlo Park, CA: Addison-Wesley.

Floridi, L. (2010). *Information: A very short introduction*. Oxford: Oxford University Press.

Fogel, J., Albert, S. M., Schnabel, F., Ditkoff, B. A., & Neugut, A. I. (2002). Use of the Internet by women with breast cancer. *Journal of Medical Internet Research, 4*(2).

Fonseca, P. A. A., & de Lima Alves, V. (2017). Cultura do Estupro: Uma análise de conteúdo sobre a percepção dos usuários via Twitter. *Idealogando: Revista de Ciências Sociais da UFPE-ISSN 2526-3552, 1*(1), 75–84.

Ford, N. (2015). *Introduction to information behaviour*. London: Facet Publishing.

Foster, A., & Ford, N. (2003). Serendipity and information seeking: An empirical study. *Journal of Documentation, 59*(3), 321–340.

Fox, S., & Duggan, M. Health. (2013). *Health Online 2013.* Washington, DC: Pew Reserach Center's Internet & American Life Project.

Fox, S., & Rainie L. (2000). *The online health care revolution: How the web helps Americans take better care of themselves.* Wasington DC: Pew Charitable Trusts.

Freelon, D. G. (2010). ReCal: Intercoder Reliability Calculation as a Web Service. *International Journal of Internet Science 5*(1), 20–33. http://www.ijis.net/ijis5_1/ijis5_1_freelon.pdf.

Ḟreeman, M., & Motsei, M. (1992). Planning health care in South Africa—is there a role for traditional healers?. *Social Science & Medicine*, *34*(11), 1183-1190.

Frieden, T. R., Damon, I., Bell, B. P., Kenyon, T., & Nichol, S. (2014). Ebola 2014—new challenges, new global response and responsibility. *New England Journal of Medicine, 371*(13), 1177–1180.

Fu, K. W., Liang, H., Saroha, N., Tse, Z. T. H., Ip, P., & Fung, I. C. H. (2016). How people react to Zika virus outbreaks on Twitter? A computational content analysis. *American journal of infection control*, *44*(12), 1700-1702.

Gaffney, D., & Puschmann, C. (2014). Data Collection on Twitter. In S. Jones (Eds.), *Twitter and Society* (pp.55–67). New York, NY: Peter Lang. l

Garland, D. (2008). On the concept of moral panic. *Crime, Media, Culture, 4*(1), 9–30.

Gerbaudo, P. (2012). *Tweets and the streets: Social media and contemporary activism*. London: Pluto Press.

Gerlitz, C., & Rieder, B. (2013). Mining one percent of Twitter: Collections, baselines, sampling. *M/C Journal, 16*(2).

Ghahremanlou, L., Sherchan, W., & Thom, J. A. (2015). Geotagging Twitter Messages in Crisis Management. *The Computer Journal, 58*(9), 1937–1954. doi: 10.1093/comjnl/bxu034

Glanz, K., Rimer, B. K., & Lewis, F.M. (Eds). (2002). *Health Behaviour and Health Education: Theory, Research and Practice* (3rd ed.) San Francisco: Jossey-Bass.

GNIP (2017). Tweet Metadata Timeline. Retrieved 9 July, 2017, from http://support.gnip.com/articles/tweet-timeline.html.

Goertzel, T. (1994). Belief in Conspiracy Theories. *Political Psychology, 15*(4), 731-742. doi: 10.2307/3791630.

González-Bailón, S., Wang, N., Rivero, A., Borge-Holthoefer, J., & Moreno, Y. (2014). Assessing the bias in samples of large online networks. *Social Networks, 38*, 16–27. doi: 10.1016/j.socnet.2014.01.004.

Google Zeitgeist. (2009). Google. Retrieved 9 July, 2017, from https://archive.google.com/intl/en/press/zeitgeist2009/.

Google. (2017). Google Trends. Retrieved 14 May, 2017, from https://trends.google.co.uk/trends/.

Graham, M. E. (2002). Health beliefs and self-breast examination in black women. *Journal of Cultural Diversity, 9*(2), 49–54.

Gray, S. M. (2003). Looking for Information: A Survey of Research on Information Seeking, Needs, and behaviour. *Journal of the Medical Library Association, 91*(2), 259.

Groshek, J. (2014). Twitter Collection and Analysis Toolkit (TCAT) at Boston University. Retrieved 05 May 2017 from http://www.bu.edu/com/bu-tcat/.

Greene, J. A., Choudhry, N. K., Kilabuk, E., & Shrank, W. H. (2011). Online social networking by patients with diabetes: A qualitative evaluation of communication with Facebook. *Journal of General Internal Medicine, 26*(3), 287–292. doi: 10.1007/s11606-010-1526-3.

Guttman, L. (1945). A basis for analyzing test-retest reliability. *Psychometrika*, 10(4), 255-282.

Guillaume, L. R. (2006). *The measles, mumps and rubella (MMR) vaccine scare: The information needs and information sources of parents.* (Unpublished doctoral thesis). University of Sheffield, Sheffield, UK.

Guillemin, J. (2006). Scientists and the history of biological weapons: A brief historical overview of the development of biological weapons in the twentieth century. *EMBO Reports, 7*(1S), S45–S49. doi: 10.1038/sj.embor.7400689.

Gulland, A. (2016). Ebola outbreak in west Africa is officially over. *BMJ: British Medical Journal (Online), 352*.

Gupta, A., Joshi, A., & Kumaraguru, P. (2012). *Identifying and Characterizing User Communities on Twitter During Crisis Events*. *Proceedings of the 2012 Workshop on Data-Driven User Behavioral Modelling and Mining from Social Media* (pp. 23–26). New York: ACM Press. doi: 10.1145/2390131.2390142.

Haensch, S., Bianucci, R., Signoli, M., Rajerison, M., Schultz, M., Kacki, S., ... Bramanti, B., (2010). Besansky, Nora J, ed. Distinct clones of Yersinia pestis caused the black death. *PLoS Pathog. 6*(10): p.e1001134. doi: 10.1371/journal.ppat.1001134.

Hansen, D., Shneiderman, B., & Smith, M. A. (2010). *Analyzing social media networks with NodeXL: Insights from a connected world*. Burlington, Massachusetts: Morgan Kaufmann.

Harris, J. K., Moreland-Russell, S., Tabak, R. G., Ruhr, L. R., & Maier, R. C. (2014). Communication about childhood obesity on Twitter. *American Journal of Public Health, 104*(7), 62–9. doi: 10.2105/AJPH.2013.301860.

Harrison, J. A., Mullen, P D and Green, L W, (1992). A meta-analysis of studies of the Health Belief Model with adults. *Health Education Research, 7*, 107–116.

Hart, C. (1998). *Doing a literature review: Releasing the social science research imagination*. London: Sage Publications.

Hartley, J. F. (1994). Case studies in organizational research. In C. Cassell & G. Symon (Eds.), *Qualitative methods in organizational research: A practical guide*, (pp. 209–29). London: Sage.

Hawksey, M. (n.d.). TAGS. Retrieved 23 December, 2014, from: https://tags.hawksey.info/.

Hay, S. I., George, D. B., Moyes, C. L., & Brownstein, J. S. (2013). Big data opportunities for global infectious disease surveillance. *PLoS medicine, 10*(4), p.e1001413.

Hayden, J. (2014). *Introduction to health behaviour theory.* Burlington, MA: Jones & Bartlett Learning.

Health Protection Agency [HPA]. (2011). [Patient Information Leaflet (PIL)]. Available http://webarchive.nationalarchives.gov.uk/20140714113442/http://www.hpa.org.uk/webc/HPAwebFile/HPAweb_C/1194947368002.

Hewis, J. (2015). Do MRI Patients Tweet? Thematic Analysis of Patient Tweets About Their MRI Experience. *Journal of Medical Imaging and Radiation Sciences. 46*(4), 396–402.

Hilton, S. and Smith, E. (2010). Public views of the UK media and government reaction to the 2009 swine flu pandemic. *BMC Public Health, 10*(1), 697.

Hochbaum, G.M. (1958). *Public Participation in Medical Screening Programs: A Socio-psychological Study* (Public Health Service Publication No. 572). Washington, DC: Government Printing Office.

Hoffman, H. (2009, April 29). Facebook maps the Swine Flu hysteria. *CNET.* Retrieved from https://www.cnet.com/news/facebook-maps-the-swine-Flu-hysteria/.

Honigsbaum, M. (2009, April 30). Swine flu and the lessons of 1918. *The Guardian.* Retrieved from http://www.theguardian.com/global/2009/apr/30/swine-flu-historical-lessons.

Howard, P. N., Duffy, A., Freelon, D., Hussain, M. M., Mari, W., & Maziad, M. (2011). *Opening closed regimes: What was the role of social media during the Arab Spring? Project on Information Technology and Political Islam Data Memo 2011.1*. Seattle: University of Washington.

Hu, Y. (2015). Health communication research in the digital age: A systematic review. *Journal of Communication in Healthcare, 8*(4), 260–288.

Ioos, S., Mallet, H. P., Goffart, I. L., Gauthier, V., Cardoso, T., & Herida, M. (2014). Current Zika virus epidemiology and recent epidemics. Medecine et maladies infectieuses, 44(7), 302–307.

Irony, n. (2017). In: *Cambridge Dictionary*. Retrieved from http://dictionary.cambridge.org/dictionary/english/irony.

Iyioha, I. O., & Nwabueze, R. N. (2016). *Comparative health law and policy: critical perspectives on Nigerian and global health law*. Routledge.

Kindhauser, M. K., Allen, T., Frank, V., Santhana, R. S., & Dye, C. (2016). Zika: the origin and spread of a mosquito-borne virus. *Bulletin of the World Health Organization*, *94*(9), 675-686C.

Crowdresearch. (2013). Is the Sample Good Enough? Comparing Twitter's Streaming API with Twitter's Firehose [Blog post]. Retrieved 1 September, 2015, from http://crowdresearch.org/blog/?p=6596.

Mason, J. (2012), *Qualitative Researching*, (2nd ed.). London, Thousand Oaks, New Delhi: Sage Publications.

Jacomy, M., Venturini, T., Heymann, S., & Bastian, M. (2014). ForceAtlas2, a continuous graph layout algorithm for handy network visualization designed for the Gephi software. *PLoS One, 9*(6), e98679.

Jansen, B. (2014, October 4). Five U.S. airports to enhance screenings for Ebola. *USA Today*. Retrieved from https://www.usatoday.com/story/news/nation/2014/10/08/ebola-travel-airport-screenings/16914661/.

Janz, N. K., Becker, M. H. (1984). The Health Belief Model: A Decade Later. *Health Education Behavior 11*(1), 1–47.

Java, A., Song, X., Finin, T., & Tseng, B. (2007). *Why we twitter: Understanding microblogging usage and communities. Proceedings of the 9th WebKDD and 1st SNA-KDD 2007 workshop on Web mining and social network analysis, August 2007* (pp. 56–65). New York: ACM Press.

Jeffares, S. (2014). Interpreting hashtag politics: Policy ideas in an era of social media. Springer, 2014.

Jin, F., Wang, W., Zhao, L., Dougherty, E., Cao, Y., Lu, C.T., & Ramakrishnan, N. (2014), Misinformation Propagation in the Age of Twitter. *Computer, 47*(12), 90–94.

Joffe, H., & Yardley, L. (2004). Content and thematic analysis. *Research methods for clinical and health psychology, 56*, 68.

Jones, J., & Salathe, M., (2009) Early assessment of anxiety and behavioural response to novel swine-origin influenza A(H1N1). *PLoS ONE 4*: p.e8032.J.

Kalyanam, J., Velupillai, S., Doan, S., Conway, M., & Lanckriet, G. (2015). Facts and Fabrications about Ebola: A Twitter Based Study. Computers and Society. *arXiv preprint arXiv:1508.02079*. Retrieved from http://arxiv.org/abs/1508.02079.

Kaplan, A. M., & Haenlein, M. (2010). Users of the world, unite! The challenges and opportunities of Social Media. *Business Horizons, 53*(1), 59–68.

Karanja, B., & Flanagan, M. (2012). How Africa tweets.

Kendra, R. L., Karki, S., Eickholt, J. L., & Gandy, L. (2015). Characterizing the Discussion of Antibiotics in the Twittersphere: What is the Bigger Picture? *J Med Internet Res, 17*(6), 154

Klout, (n.d). Retrieved 19 June, 2017, from https://klout.com/home.

Kouloumpis, E., Wilson, T., & Moore, J. D. (2011). Twitter sentiment analysis: The good the bad and the omg!. *Icwsm, 11*(538–541), 164.

Krippendorff, K. (2004). Content Analysis: An Introduction to Its Methodology. *Education* (79). doi: 10.2307/2288384.

Lachlan, K. A., Spence, P. R., Lin, X., Najarian, K., & Del Greco, M. (2015). Social media and crisis management: CERC, search strategies, and Twitter content. *Computers in Human Behavior, 54*, 647–652. doi: 10.1016/j.chb.2015.05.027

Lazard, A. J., Scheinfeld, E., Bernhardt, J. M., Wilcox, G. B., & Suran, M. (2015). Detecting themes of public concern: A text mining analysis of the Centers for Disease Control and Prevention's Ebola live Twitter chat. *American Journal of Infection Control, 43*(10), 1109–1111. doi: 10.1016/j.ajic.2015.05.025

Lazer, D., Kennedy, R., King, G., & Vespignani, A. (2014). The parable of Google Flu: Traps in big data analysis*. Science, 343*(6176), 1203–1205.

Leander, K. M., & McKim, K. K. (2003). Tracing the everyday'sitings' of adolescents on the internet: A strategic adaptation of ethnography across online and offline spaces. *Education, communication & information*, *3*(2), 211-240.

Lee, J. L., DeCamp, M., Dredze, M., Chisolm, M. S., & Berger, Z. D. (2014). What are health-related users tweeting? A qualitative content analysis of health-related users and their messages on twitter. *Journal of medical Internet research, 16*(10).

Lewin K. (1951). *Field theory in social science: Selected theoretical papers.* New York: Harper & Row.

Li, Y., Wang, X., Lin, X., & Hajli, M. (2016). Seeking and sharing health information on social media: A net valence model and cross-cultural comparison*. Technological Forecasting and Social Change, 4*.

Litwin, M. S., & Fink, A. (Eds.) (1995). *How to Measure Survey Reliability and Validity, Vol. 7.: The Survey Kit.* Thousand Oaks, CA: Sage Publications.

NVivo. (2012). Qualitative data analysis Software, Version 10. QSR International Pty Ltd. [Qualitative data analysis software].

Liu, Y., Mei, Q., Hanauer, D. A., Zheng, K., Lee, J. M. (2016). Use of Social Media in the Diabetes Community: An Exploratory Analysis of Diabetes-Related Tweets. *JMIR Diabetes, 1*(2). doi: b10.2196/diabetes.6256.

Loader, B. D., & Mercea, D. (2011). Networking democracy? Social media innovations and participatory politics. *Information, Communication & Society, 14*(6), 757–769.

Lotan, G., Graeff, E., Ananny, M., Gaffney, D., Pearce, I., & Boyd, D. (2011). The Revolutions Were Tweeted: Information Flows During the 2011 Tunisian and Egyptian Revolutions. *International Journal of Communication,* 5, 1375–1405.

Mackenzie, N. & Knipe, S. (2006). Research dilemmas: Paradigms, methods and methodology. *Educational Research, 16*(2), 193–205.

Mangili, A. & Gendreau, M.A. (2005). Transmission of infectious diseases during commercial air travel. *The Lancet*, *365*(9463), 989–996.

Manguvo, A., & Mafuvadze, B. (2015). The impact of traditional and religious practices on the spread of Ebola in West Africa: Time for a strategic shift. *The Pan African Medical Journal, 22*(S1), 9. doi: 10.11694/pamj.supp.2015.22.1.6190.

Marchionini, G. (1997). *Information seeking in electronic environments (No. 9).* Cambridge: Cambridge University Press.

Marwick, A. (2013). Ethnographic and Qualitative Research on Twitter. In: K. Weller, A. Bruns, C. Puschmann, J. Burgess, J. & M. Mahrt (Eds.), *Twitter and Society*, (pp. 109–122). New York: Peter Lang.

Marwick, A., & Boyd, D. (2014). Networked privacy: How teenagers negotiate context in social media. *New Media & Society, 16*, 1051–1067. doi: 10.1177/1461444814543995.

Maslow, A., & Lewis, K. J. (1987). *Maslow's hierarchy of needs* (pp. 14). New York: Salenger Incorporated.

Matson, J. L., Wilkins, J., & Gonzalez, M. (2008). Early identification and diagnosis in autism spectrum disorders in young children and infants: How early is too early*? Research in Autism Spectrum Disorders, 2*, 75–84.

Line, M. B. (1974). Draft definitions: Information and library needs, wants, demands and uses. *Aslib Journal of Information Management, 26*(2), 87.

McCandless, D. (2009). *Information is beautiful* (pp. 1–255). London: Collins.

McHugh, M. L. (2012). Interrater reliability: The kappa statistic. *Biochemia medica, 22*(3), 276–282.

McKee, R. (2013). Ethical issues in using social media for health and health care Research. *Health Policy, 110.2*, 298–301.

Mear, C. (2016). *An international NGO startup's use of social media technology: The case of Anti Corruption International: A discursive analysis on the organizational use of the term 'corruption' on Twitter*. (Unpublished Masters thesis). Norwegian University of Science and Technology, Trondheim, Norway.

Mendoza, M., Poblete, B., & Castillo, C. (2010). *Twitter under crisis. Proceedings of the First Workshop on Social Media Analytics – SOMA '10* (pp. 71–79). New York: ACM Press. doi: 10.1145/1964858.1964869.

Mertens, D.M. (2005). *Research methods in education and psychology: Integrating diversity with quantitative and qualitative approaches* (2nd Ed.). Thousand Oaks: Sage.

Mesch, G. S., Schwirian, K. P., & Kolobov, T. (2013). Attention to the media and worry over becoming infected: the case of the Swine Flu (H1N1) Epidemic of 2009. *Sociology of health & illness, 35*(2), 325–331.

Miles, M. B., Huberman, A. M., & Saldana, J (2013). *Qualitative data analysis*. New York: Sage.

Mislove, A., Lehmann, S., Ahn, Y., Onnela, J., & Rosenquist, J. N. (2011). Understanding the Demographics of Twitter Users. *Artificial Intelligence,* 554–557. Retrieved from http://www.aaai.org/ocs/index.php/ICWSM/ICWSM11/paper/viewFile/2816/3234.

Mlakar, J., Korva, M., Tul, N., Popović, M., Poljšak-Prijatelj, M., Mraz, J., ... & Vizjak, A. (2016). Zika virus associated with microcephaly. *N Engl J Med*, *2016*(374), 951-958.

Mocanu, D., Rossi, L., Zhang, Q., Karsai, M., & Quattrociocchi, W. (2015). Collective attention in the age of (mis) information. *Computers in Human Behavior, 51*, 1198–1204.

Moon, S., Sridhar, D., Pate, M. A., Jha, A. K., Clinton, C., Delaunay, S. ... Goosby, E. (2015). Will Ebola change the game? Ten essential reforms before the next pandemic. The report of the Harvard-

LSHTM Independent Panel on the Global Response to Ebola. *The Lancet , 386*(10009), 2204–2221.

Morahan-Martin, J. M. (2004). How internet users find, evaluate, and use online health information: A cross-cultural review. *CyberPsychology & Behavior, 7*(5), 497–510.

Morgan, D. L. (2007). Paradigms Lost and Pragmatism Regained: Methodological Implications of Combining Qualitative and Quantitative Methods. *Journal of Mixed Methods Research, 1*(1), 48–76. doi: 10.1177/2345678906292462.

Morstatter, F., Pfeffer, J., Liu, H., & Carley, K. (2013). *Is the Sample Good Enough? Comparing Data from {Twitter's} Streaming {API} with {Twitter's} {Firehose}*. *Proceedings of ICWSM*. California: AAAI.

Mowbray, M. (2014). Automated Twitter accounts. In: S. Jones, (Eds.), *Twitter and Society* (pp.183–191). New York, NY: Peter Lang.

Mozdeh (n.d.). Mozdeh Twitter Time Series Analysis. Retrieved 23 December, 2014, from http://mozdeh.wlv.ac.uk/.

Murray, C. J. L., & Lopez, A. D. (1996). *The global burden of disease: A comprehensive assessment of mortality and disability from diseases*. Geneva, Switzerland: World Health Organization.

Myslín, M., Zhu, S.-H., Chapman, W., & Conway, M. (2013). Using twitter to examine smoking behaviour and perceptions of emerging tobacco products. *Journal of Medical Internet Research, 15*(8), 174. doi: 10.2196/jmir.2534.

National Cancer Institute (2003). *Theory at a Glance: A Guide for Health Promotion Practice*. Washington, DC: US Department of Health and Human Services.

Nelson, R., & Staggers, N. (2013). *Health informatics: An interprofessional approach.* Maryland Heights, Missouri: Mosby.

Ngade, I., Singer, M., Marcus, O., & Lara, J. E. H. (2016). Washing Away Ebola. *A Companion to the Anthropology of Environmental Health*, 157–172.

NHS Choices [NHS]. (2013). Norovirus. Retrieved 2 December, 2012, from http://www.nhs.uk/Conditions/Norovirus/Pages/Introduction.aspx.

NHS Choices [NHS]. (2015). Swine flu (H1N1). Retrieved 20 October, 2015, from http://www.nhs.uk/conditions/pandemic-flu/Pages/Introduction.aspx.

Nicholas, D. (2003). *Assessing information needs: Tools, techniques and concepts for the internet age* (2nd Edn.). London & New York: Routledge.

Nicholas, D., & Herman, E. (2010). *Assessing Information Needs in the Age of the Digital Consumer* (3rd Edn.). London & New York: Routledge.

Nicolson, J (2009, November 11). The war was over - but Spanish Flu would kill millions more. *The Telegraph.* Retreived from http://www.telegraph.co.uk/health/6542203/The-war-was-over-but-Spanish-Flu-would-kill-millions-more.html.

Niswander, K. R. (1965). Medical abortion practices in the United States. *W. Res. L. Rev.*, *17*, 403.

Neuendorf, K. A. (2002). *The content analysis guidebook*. Thousand Oaks , CA: Sage.

O'Connor, B., & Balasubramanyan, R. (2010). From *tweets to polls: Linking text sentiment to public opinion time series. Proceedings of the Fourth International AAAI Conference on Weblogs and Social Media* (pp. 122–129). Washington, DC: AAAI.

Odlum, M., & Yoon, S. (2015). What can we learn about the Ebola outbreak from tweets? *American Journal of Infection Control, 43*(6), 563–71. doi: 10.1016/j.ajic.2015.02.023.

Olumwafemi, O.S., Elia, G., Rolf, W. (2014). Ebola, Twitter, and misinformation: a dangerous combination? British Medical Journal; 349 ;g6178.

Opsahl, T., Agneessens, F., & Skvoretz, J. (2010). Node centrality in weighted networks: Generalizing degree and shortest paths. *Social networks, 32*(3), 245–251.

Pak, A., & Paroubek, P. (2010, May). Twitter as a corpus for sentiment analysis and opinion mining. In *LREc* (Vol. 10, No. 2010).

Pang, B., & Lee, L. (2008). Opinion Mining and Sentiment Analysis. *Foundations and Trends® in Information Retrieval*, *2*(1–2), pp.1–135. doi: 10.1561/1500000011.

Panjwani, A., & Wilson, A. (2016). What Is Stopping the Use of Genetically Modified Insects for Disease Control?. *PLoS pathogens*, *12*(10), e1005830.

Parkinson, J. H. (2016, November 3). Twitter is replacing favourites with likes – but does anyone heart it? *The Guardian.* Retrieved from https://www.theguardian.com/technology/2015/nov/03/twitter-replacing-favourites-with-likes-does-anyone-heart.

Paulus, T., Woods, M., Atkins, D. P., & Macklin, R. (2017). The discourse of QDAS: Reporting practices of ATLAS. ti and NVivo users with implications for best practices. *International Journal of Social Research Methodology, 20*(1), 35–47.

Paz, S., & Semenza, J. C. (2016). El Niño and climate change—contributing factors in the dispersal of Zika virus in the Americas?. *The Lancet, 387*(10020), 745.

Pennacchiotti, M., & Popescu, A. M. (2011). A Machine Learning Approach to Twitter User Classification. *ICWSM, 11*(1), 281–288.

Perrett, M. (2009, April 28th). Swine Flu: TUI and Thomas Cook Cancel Mexico flights. *Travel Weekly*. Retrieved from http://www.travelweekly.co.uk/articles/30850/swine-Flu-tui-and-thomas-cook-cancel-mexico-flights.

Petrovic, S., Osborne, M., Mccreadie, R., Macdonald, C., & Ounis, I. (2013). *Can twitter replace newswire for breaking news? Proceedings of ICWSM*. California: AAAI.

Pew Research Center (2016). Social Media Update 2016. Retrieved 9 March, 2007, from http://assets.pewresearch.org/wp-content/uploads/sites/14/2016/11/10132827/PI_2016.11.11_Social-Media-Update_FINAL.pdf**.**

Phuvipadawat, S., & Murata, T. (2010). *Breaking news detection and tracking in Twitter. In Proceedings Web Intelligence and Intelligent Agent Technology (WI-IAT), 2010. IEEE/WIC/ACM International Conference 3*, 120–123: IEEE.

Pickard, A. J. (Ed.) (2007). *Research Methods in Information.* London: Facet Publishing.

Pilkington, E., McGreal, C., & Rice, X. (2009, May 1). President Barack Obama guard on Mexico trip 'has swine Flu'. *The Guardian.* Retrieved from https://www.theguardian.com/world/2009/may/01/obama-swine-Flu-mexico.

Poole, H. (1985). *Theories of the middle range.* Norwood, NJ: Ablex.

Pozzi, F. A., Fersini, E., Messina, E., & Liu, B. (2016). *Sentiment Analysis in Social Networks*. Burlington, Massachusetts: Morgan Kaufmann.

Prieto, V. M., Matos, S., Álvarez, M., Cacheda, F., & Oliveira, J. L. (2014). Twitter: A good place to detect health conditions. *PloS ONE, 9*(1), e86191. doi: 10.1371/journal.pone.0086191.

Procter, R., Vis, F., & Voss, A. (2013). Reading the riots on Twitter: Methodological innovation for the analysis of big data. *International Journal of Social Research Methodology, 16*(3), 197–214. doi: 10.1080/13645579.2013.774172.

Punch, K. F. (2013). *Introduction to social research: Quantitative and qualitative approaches*. New York: Sage.

Purohit, H., Hampton, A., Shalin, V. L., Sheth, A. P., Flach, J., & Bhatt, S. (2013). What kind of #conversation is Twitter? Mining #psycholinguistic cues for emergency coordination. *Computers in Human Behavior, 29*(6), 2438–2447. doi: 10.1016/j.chb.2013.05.007.

R.L. Heath, L.X. Blonder (2003). Conversational humour among stroke survivors. *Humor, 16*(1), 91–106.

Randolph, W., & Viswanath, K. (2004). Lessons learned from public health mass media campaigns: Marketing health in a crowded media world. *Annual Review of Public Health, 25*, 419–437. doi: 10.1146/annurev.publhealth.25.101802.123046.

Ragin, C. C. (2014). *The comparative method: Moving beyond qualitative and quantitative strategies*. Univ of California Press.

Reeves S, Albert M, Kuper A, Hodges B D. (2008). Why use theories in qualitative research? *BMJ, 337*, 949.

Reyneke, M., Pitt, L., & Berthon, P. R. (2011). Luxury wine brand visibility in social media: An exploratory study. *International Journal of Wine Business Research, 23*(1), 21–35.

Roa, M. (2016). Zika virus outbreak: reproductive health and rights in Latin America. *The Lancet*,

387(10021), 843.

Robillard, J. M., Johnson, T. W., Hennessey, C., Beattie, B. L., & Illes, J. (2013). Aging 2.0: Health information about dementia on Twitter. *PloS ONE, 8*(7), p.e69861. doi: 10.1371/journal.pone.0069861.

Robson, C. (2002). Real World Research: A Resource for Social Scientists and Practitioner-Researchers. booksgooglecom (Vol. 2nd). New York: John Wiley and Sons. Retrieved from http://books.google.com/books?id=DkplMcAysFQC\nhttp://media.matthewsbooks.com.s3.amazonaws.com/documents/tocwork/063/9780631213055.pdf.

Rogers, R. W., (1983). Cognitive and physiological processes in fear appeals and attitude change: A revised theory of protection motivation. In: J. T. Cacioppo, R. E. Petty & D. Shapiro (Eds.), *Social Psychophysiology: A Sourcebook* (pp. 153–176). New York, Guilford Press.

Rogers, R. W. (1975). A protection motivation theory of fear appeals and attitude change1. *The Journal of Psychology, 91*(1), 93–114.

Roos, D. (2015). What is an API? – HowStuffWorks. *HowStuffWorks*. Retrieved 12 March., 2015, from http://money.howstuffworks.com/business-communications/how-to-leverage-an-api-for-conferencing1.htm.

Rosenblum, J. D. (2010). "Why Do You Refuse to Eat Pork?": Jews, Food, and Identity in Roman Palestine*. Jewish Quarterly Review, 100*(1), 95–110.

Russell, S., Middleton-Green, L., & Johnston, B. (2015). Using social media to create discussion. *International Journal of Palliative Nursing, 21*(11), 525–526.

Rutsaert, P., Pieniak, Z., Regan, A., McConnon, A., & Verbeke, W. (2013). Consumer interest in receiving information through social media about the risks of pesticide residues. *Food Control, 34*, 386–392. doi: 10.1016/j.foodcont.2013.04.030.

Rutsaert, P., Pieniak, Z., Regan, A., McConnon, A., Kuttschreuter, M., Lores, M., & Verbeke, W. (2014). Social media as a useful tool in food risk and benefit communication? A strategic orientation approach. *Food Policy, 46*, 84–93. doi: 10.1016/j.foodpol.2014.02.003.

Sakaki, T., Okazaki, M., & Matsuo, Y. (2010). *Earthquake shakes Twitter users. Proceedings of the 19th International Conference on World Wide Web* (pp. 851). New York: ACM Press.

Salathé, M., & Khandelwal, S. (2011). Assessing vaccination sentiments with online social media: Implications for infectious disease dynamics and control. *PLoS Computational Biology, 7*(10), p.e1002199. doi: 10.1371/journal.pcbi.1002199.

Santos, J. C., & Matos, S. (2014). Analysing Twitter and web queries for flu trend prediction. *Theoretical Biology & Medical Modelling, 11*(1), S6. doi: 10.1186/1742-4682-11-S1-S6.

Sarcasm (2017). In: *Cambridge Dictionary*. Retrieved from http://dictionary.cambridge.org.

Saunders, M., Lewis, P., & Thornhill, A. (2009). *Research Methods for Business Students*. New Jersey: Financial Times Prentice Hall. Retrieved from https://books.google.com/books?id=u-txtfaCFiEC&pgis=1.

Savolainen, R. (2006). Time as a Context of Information Seeking. *Library and Information Science Research, 28*(1), 110–27.

Savolainen, R. (2009). Information use and information processing: Comparison of conceptualizations. *Journal of Documentations, 65*(2), 187–207.

Scanfeld, D., Scanfeld, V., & Larson, E. L. (2010). Dissemination of health information through social networks: Twitter and antibiotics. *American Journal of Infection Control, 38*(3), 182–8. doi: 10.1016/j.ajic.2009.11.004.

Schement, J. R ., & Reuban, B.D. (1993). *Between communication and information.* New Brunswick, NJ: Transaction Publishers.

Schnirring, L. (2013, July 11). CDC says norovirus sickens up to 21 million a year. *CIDRAP*. Retrieved from http://www.cidrap.umn.edu/news-perspective/2013/07/cdc-says-norovirus-sickens-21-million-year.

Seale, H., Dwyer, D. E., Cowling, B. J., Wang, Q., Yang, P. & MacIntyre, C. R. (2009). A review of medical masks and respirators for use during an influenza pandemic. *Influenza and Other Respiratory Viruses, 3*, 205–206. doi: 10.1111/j.1750-2659.2009.00101.x.

Signorini A., Segre A. M., Polgreen P. M. (2011) The Use of Twitter to Track Levels of Disease Activity and Public Concern in the U.S. during the Influenza A H1N1 Pandemic. *PLoS ONE 6*(5), p.e19467. doi: 10.1371/journal.pone.0019467.

Simon, H. A. (1956). Rational choice and the structure of the environment. *Psychological review, 63*(2), 129.

Simon, T., Goldberg, A., Aharonson-Daniel, L., Leykin, D., & Adini, B. (2014). Twitter in the cross fire – the use of social media in the Westgate Mall terror attack in Kenya. *PloS ONE, 9*(8), p.e104136. doi: 10.1371/journal.pone.0104136

Simonsen, L., Spreeuwenberg, P., Lustig, R., Taylor, R. J., Fleming, D. M., Kroneman, M., ... Paget, W. J. (2013). Global mortality estimates for the 2009 Influenza Pandemic from the GLaMOR project: A modeling study. *PLoS medicine, 10*(11), p.e1001558.

Sinnenberg, L., Buttenheim, A. M., Padrez, K., Mancheno, C., Ungar, L., & Merchant, R. M. (2017). Twitter as a tool for health research: A systematic review. *American Journal of Public Health (ajph).*

Sloan, L. (2017). Who Tweets in the United Kingdom? Profiling the Twitter Population Using the British Social Attitudes Survey 2015. *Social Media+ Society, 3*(1), 2056305117698981.

Smailhodzic, E., Hooijsma, W., Boonstra, A., & Langley, D. J. (2016). Social media use in healthcare: A systematic review of effects on patients and on their relationship with healthcare professionals. *BMC Health Services Research, 16*(1), 442. doi: 10.1186/s12913-016-1691-0.

Smith, R. D., & Sommers, T. (2003). *Assessing the economic impact of public health emergencies of international concern: the case of SARS*. Geneva: World Health Organization.

Smith, R. D. (2006). Responding to global infectious disease outbreaks: lessons from SARS on the role of risk perception, communication and management. *Social Science & Medicine, 63*(12), 3113–3123.

Srivastava, A., & Thomson, S. B. (2009). Framework analysis: A qualitative methodology for applied policy research. *JOAGG, 4*(2).

Starnes, D. S., Yates, D., & Moore, D. S. (2010). *The practice of statistics*. London: Macmillan.

Steckler, A., McLeroy, K. R., Goodman, R. M., Bird, S. T., & McCormick, L. (1992). Toward integrating qualitative and quantitative methods: An introduction. *Health Education Quarterly, 19*(1), 1–8.

Stefanidis, A., Vraga, E., Lamprianidis, G., Radzikowski, J., Delamater, P. L., Jacobsen, K. H., ... Crooks, A.(2017). Zika in Twitter: Temporal Variations of Locations, Actors, and Concepts. *JMIR Public Health Surveill, 3*(2), 22. doi: 10.2196/publichealth.6925. PMID: 28428164.

Steinbock, D. (n.d.). TAG Crowd. Retrieved 10 April, 2017, from http://tagcrowd.com/.

Stoppler, C. M. (2016). Mad Cow Disease in Humans: Facts, Symptoms, Causes & Treatment. *eMedicineHealth*. Retrieved from http://www.emedicinehealth.com/mad_cow_disease_and_variant_creutzfeldt-jakob/article_em.htm.

Storr, W. (2014, November 8). 'Ebola is man-made', and other crazy conspiracy theories. *The Telegraph*. Retrieved from http://www.telegraph.co.uk/culture/books/11216881/Ebola-is-man-made-and-other-crazy-conspiracy-theories.html.

Stretcher, V., & Rosenstock, I. M. (1997). The Health Belief Model. In: K. Glanz, F. M. Lewis, & B. K. Rimer (Eds.), *Health Behaviour and Health Education: Theory, Research and Practice* (2nd ed.). San Francisco: Jossey-Bass.

Strickland, J., & Chandler, N. (2015). Twitter's API – HowStuffWorks. *HowStuffWorks*. Retrieved 12 March, 2015, from http://computer.howstuffworks.com/internet/social-networking/networks/twitter2.htm.

Sueki, H. (2014). The association of suicide-related Twitter use with suicidal behaviour: A cross-sectional study of young internet users in Japan. *Journal of Affective Disorders, 170C*, 155–160. doi: 10.1016/j.jad.2014.08.047.

Sullivan, M. (2015, February 4). Twitter CEO Dick Costolo says trolls are driving away 'core' users. *VentureBeat.* Retrieved from: http://venturebeat.com/2015/02/04/twitter-ceo-dick-costolo-says-trolls-are-driving-away-core-users/?utm_campaign=naytev&utm_content=54d374b6e4b093d618544d80.

Sutter, D. J. (2009, April 30). Swine Flu creates controversy on Twitter. *CNN*. Retrieved from http://edition.cnn.com/2009/TECH/04/27/swine.flu.twitter/.

Sutton, S. (2002). Health Behavior: Psychosocial Theories. *Social Cognition*, 6499–6506.

Szomszor, M., & Kostkova, P. (Eds.). (2014). Lecture Notes of the Institute for Computer Sciences, Social-Informatics and Telecommunications Engineering. *Springer, 69*, 18–26.

Szomszor, M., Kostkova, P., & St Louis, C. (2011). *Twitter informatics: Tracking and understanding public reaction during the 2009 Swine Flu pandemic. Proceedings of the 2011 IEEE/WIC/ACM International Conference on Web Intelligence, WI 2011* (Vol. 1, pp. 320–323). New York: ACM Press. doi:10.1109/WI-IAT.2011.311.

Szomszor, M., Kostkova, P., & de Quincey, E. (2010). *#Swineflu: Twitter predicts swine flu outbreak in 2009. Proceedings of the 3rd International Conference on Electronic Healthcare, M* (pp. 13–15), Dec *2010*, Casablanca, Morocco: Springer. doi:10.1007/978-3-642-23635-8_3.

Tamsin, R. (2013, December 11). '#Barf – How Twitter Can Reduce The Spread Of Norovirus'. *The Guardian*. Retrieved from http://www.theguardian.com/public-leaders-network/2013/dec/11/twitter-reduce-spread-norovirus.

Tanis, M. (2008). Health-related on-line forums: what's the big attraction? *Journal of Health Communication*, 13(7), 698–714.

Taubenberger, J. K., Morens, D. M. (2006). 1918 influenza: The mother of all pandemics. *Emerging Infectious Diseases, 12,* 15–22. Retrieved from http://wwwnc.cdc.gov/eid/article/12/1/05-0979_article.htm#suggestedcitation**.**

Taylor, L. H., Latham, S. M., & Mark, E. J. (2001). Risk factors for human disease emergence. *Philosophical Transactions of the Royal Society of London B: Biological Sciences, 356*(1411), 983–989.

Taylor-Smith, E. (2016). *Investigating oral presentation skills and non-verbal communication techniques in UAE classrooms: A thesis in teaching English to speakers of other languages.* (Unpublished doctoral thesis). American University of Sharjah, Sharjah, United Arab Emirates.

Teasdale, E., & Yardley, L. (2011) Understanding responses to government health recommendations: Public perceptions of government advice for managing the H1N1 (swine flu) influenza pandemic. *Patient Education and Counseling, 85*(3), 413–418.

Telegraph (2017, November 7). Twitter is letting everyone post 280 character tweets. *The Telegraph*. Retrieved from http://www.telegraph.co.uk/technology/2017/11/07/twitter-letting-everyone-post-280-character-tweets/

Telegraph (2016, February 2). Zika latest: Airlines offer refunds to pregnant passengers. *The Telegraph*. Retrieved from http://www.telegraph.co.uk/travel/news/Zika-latest-Airlines-offer-refunds-to-pregnant-passengers/

Telegraph (2009, April 29). Swine Flu: Do Surgical Masks really work? *The Telegraph.* Retrieved from http://www.telegraph.co.uk/news/health/swine-Flu/5239580/Swine-Flu-do-surgical-masks-really-work.html.

Thackeray, R., Crookston, B. T., & West, J. H. (2013). *J Med Internet Res, 15*(1), 21.

Champion, V. L., & Sugg Skinner, C. (2008). The Health Belief Model. In: K. Glanz, F. M. Lewis, & B. K. Rimer, (Eds.). *Health Behaviour and Health Education: Theory, Research and Practice*. San Francisco: Jossey-Bass.

The University of Sheffield (2006). Ethics Policy Governing Research Involving Human Participants, Personal Data and Human Tissue. Retrieved 1 September 2017.

Thelwall, M. (2015). Evaluating the comprehensiveness of Twitter search API results: a four step method. *International Journal of Scientometrics, Informetrics and Bibliometrics*, *18*(1), 1-10.

Thelwall, M. (2014). Sentiment Analysis and Time Series with Twitter. In: S. Jones (Ed.) *Twitter and Society* (pp. 83–95). New York, NY: Peter Lang.

Timeline: 2009 swine flu outbreak. (2009). *NBC News.* Retrieved 10 April, 2017, from http://www.nbcnews.com/id/30624302/ns/health-cold_and_flu/t/timeline-swine-flu-outbreak/#.WOtL73qXyYI.

Towers, S., Afzal, S., Bernal, G., Bliss, N., Brown, S., Espinoza, B., … Castillo-Chavez, C. (2015). Mass Media and the Contagion of Fear: The Case of Ebola in America. *PLoS ONE, 10*(6), p.e0129179. doi: 10.1371/journal.pone.0129179.

Trochim, W., Donnelly, J., Arora, K. (2014). *Research Methods: The Essential Knowledge Base* (2nd edn.). California: Wadsworth Publishing Co Inc.

Tweet activity dashboard (n.d.). Twitter Help Center. Retrieved 19 August, 2017, from https://support.twitter.com/articles/20171990#.

Twitter Developers Blog (2014). [Blog post] Retrieved 23 December, 2014, from https://dev.twitter.com.

Twitter Q1 2017 Company Metrics. (2017). Number of monthly active Twitter users worldwide from 1st quarter 2010 to 1st quarter 2017 (in millions). *Statista*. Retrieved 19 June, 2017, from https://www.statista.com/statistics/282087/number-of-monthly-active-twitter-users/.

Varol, O., Ferrara, E., Davis, C., Menczer, F., & Flammini, A. (2017). Online Human-Bot Interactions: Detection, Estimation, and Characterization. In *Proceedings of the Eleventh International AAAI Conference on Web and Social Media (ICWSM 2017), 15-18 May 2017.* (pp. 280–289) Montreal, Canada: AAAI.

Velasco, E., Agheneza, T., Denecke, K., Kirchner, G., & Eckmanns, T. (2014). Social media and Internet-Based data in global systems for public health surveillance: A systematic review. *The Milbank Quarterly, 92*(1), 7–33.

Vis, F. (2013). A critical reflection on Big Data: Considering APIs, researchers and tools as data makers. *First Monday*, *18*(10).

Vis, F. (2013). Twitter as a reporting tool for breaking news. *Digital Journalism, 1*(1), 27–47. doi: 10.1080/21670811.2012.741316.

Visibrain.com (n.d.) Retrieved 03 December, 2017, from: http://www.visibrain.com/en/.

Wagner, K. (2017, March 3). Usernames no longer count towards Twitter's 140-charcter limit. *Recode*. Retrieved from https://www.recode.net/2017/3/30/15116794/twitter-longer-tweets-140-character-limit-username-reply.

Wald, R., Khoshgoftaar, T. M., Napolitano, A., & Sumner, C. (2013). Which users reply to and interact with twitter social bots?. In Proceedings Tools with Artificial Intelligence (ICTAI), November 2013 IEEE 25th International Conference (pp. 135–144). IEEE.

Wang, M. D., & Jolly, A. M. (2004). Changing virulence of the SARS virus: The epidemiological evidence. *Bulletin of the World Health Organization, 82*(7), 547–548.

Wang, T. T., & Palese, P. (2009). Unraveling the mystery of swine influenza virus. *Cell, 137*(6), 983–985.

Warren, C. (2015, July 5). Twitter's Project Lightening could be called 'Moments'. *Mashable.* Retrieved from http://mashable.com/2015/07/05/twitter-project-lightning-moments/#VZNMHYaDIsqg.

Wasserman, S., & Faust, K. (1994). *Social network analysis: Methods and applications,* Vol. 8. Cambirdge: Cambridge University Press.

Weaver, K., & Olson, J. K. (2006). Understanding paradigms used for nursing research. Journal *of Advanced Nursing, 53*(4), 459–69. doi: 10.1111/j.1365-2648.2006.03740.

Webometric Analyst (n.d.). Webometric Analyst 2.0. Retrieved 27 July, 2015, from http://lexiurl.wlv.ac.uk/.

Weitkunat, R., Pottgieber, C., Meywer, N., Crispin, A., Fischer, R., Schotten, K., ... Uveria, K. (2003). Perceived risk of bovine spongiform encephalopathy and dietary behaviour. *Journal of Health Psychology, 8*(3),373–382.

Weaver, M. (2016, February 01). Tata considers renaming new Zica car as Zika virus spreads. Retrieved December 07, 2017, from https://www.theguardian.com/business/2016/feb/01/zica-car-tata-homophone-hell-carmarker-launch-indian.

Weller, K., Bruns, A., Burgess, J. E., Merja, M.,& Puschmann, C. (2014). Twitter and society: An introduction. In: K. Weller, A. Bruns, J. Burgess, M. Mahrt & C. Puschmann (Eds.), *Twitter and Society*. New York: Peter Lang, New York.

Williamson, A. (2014, August 24). Ebola Trials: A White Man's Privilege. *Doctus Project.* Retrieved from http://www.doctusproject.com/2014/08/24/ebola-trials-white-mans-privilege/.

Williamson, L. M., Parkes, A., Wight, D., Petticrew, M., & Hart, G. J. (2009). Limits to modern contraceptive use among young women in developing countries: a systematic review of qualitative research. *Reproductive health*, *6*(1), 3.

Wilson, T. D. (1984). The cognitive approach to information seeking behaviour and information use. *Social Science Information Studies, 4*, 197–204.

World Health Organization (2009a). Current WHO phase of pandemic alert for Pandemic (H1N1) 2009. Geneva, Switzerland: World Health Organization. Available from http://www.who.int/csr/disease/swineflu/phase/en/.

World Health Organization (2009b). *Joint FAO/WHO/OIE Statement on influenza A(H1N1) and the safety of pork* [Press release]. Geneva, Switzerland: World Health Organization. Available from http://www.who.int/mediacentre/news/statements/2009/h1n1_20090430/en/.

World Health Organization. (2009c). *Pandemic H1N1 2009*. Geneva, Switzerland: World Health Organization.

World Health Organization. (2015). *Ebola virus disease.* Geneva, Switzerland: World Health Organization. Available from http://www.who.int/mediacentre/factsheets/fs103/en/.

Yahoo! Buzz Index. (2003). The Buzz Index Weekly – the top 20 searchers. Retrieved 26 August, 2017, from http://buzz.yahoo/weekly.

Yin, R. K., (1994). *Case Study Research: Design and Methods*. California: Sage Publications.

Yom-Tov, E., Borsa, D., Cox, I. J., & McKendry, R. A. (2014). Detecting disease outbreaks in mass gatherings using Internet data. *Journal of Medical Internet Research, 16*, 154. doi: 10.2196/jmir.3156.

Young, S. D., Rivers, C., & Lewis, B. (2014). Methods of using real-time social media technologies for detection and remote monitoring of HIV outcomes. *Preventive Medicine, 63*, 112–5. doi: 10.1016/j.ypmed.2014.01.024.

Zappavigna, M. (Ed.) (2012). *Discourse of Twitter and social media: How we use language to create affiliation on the Web*. New York: Continuum.

Zerbinos, E. (1990). Information seeking and information processing: Newspapers *versus* videotext. *Journalism Quarterly, 67*, 920–929.

Zhao, Z., & Mei, Q. (2013, May). *Questions about questions: An empirical analysis of information needs on Twitter. Proceedings of the 22nd international conference on World Wide Web* (pp. 1545–1556). New York: ACM Press.

Zsófia, D. (2016), Laughing at cancer: Humour, empowerment, solidarity and coping online. *Journal of Pragmatics, 101*, 18–30. doi: 10.1016/j.pragma.2016.05.010. ISSN 0378-2166.

Qsrinternational (2016). Swine Flu and Ebola: A Twitter analysis using NVivo. Retrieved 7 September, 2017, from http://www.qsrinternational.com/blog/analyzing-tweets-about-infectious-disease-outbreak.

Appendix 1 Ethical Approval

The University Of Sheffield.

Downloaded: 03/09/2017
Approved: 14/07/2016

Wasim Ahmed
Registration number: 140135011
Information School
Programme: PhD Information Studies (Social)

Dear Wasim

PROJECT TITLE: Using Twitter data to provide insights into health conditions and health-related events
APPLICATION: Reference Number 009319

On behalf of the University ethics reviewers who reviewed your project, I am pleased to inform you that on 14/07/2016 the above-named project was **approved** on ethics grounds, on the basis that you will adhere to the following documentation that you submitted for ethics review:

- University research ethics application form 009319 (dated 06/06/2016).
- Participant information sheet 1019066 version 1 (01/06/2016).
- Participant consent form 1019067 version 1 (01/06/2016).

The following optional amendments were suggested:

See amendments above - however, these are SUGGESTED amendments, and the research may proceed as described.

If during the course of the project you need to deviate significantly from the above-approved documentation please inform me since written approval will be required.

Yours sincerely

Matt Jones
Ethics Administrator
Information School

Appendix 2 Code Frame for Tweets relating to the Ebola Virus disease Outbreak

Adapted from: Chew, C., & Eysenbach, G. (2010). Pandemics in the age of Twitter: Content analysis of tweets during the 2009 H1N1 outbreak. PLOS ONE, 5(11).

1) Primary Code Frame and 2) Secondary Code Frame and explain what the coding at each of the levels aims to do. The aim with the Primary Code Frame is to 'sift' resource and spam tweets from the rest. At the second stage, the tweets which are not resource or spam tweets are subjected to further analysis

For Tweet 1, assign primary code, assign secondary code then go to tweet 2 and repeat, etc.

1 Primary Code Frame

Based on reading the tweet and examining the linked external content (URLs), assign it to one of the following categories:

1=Resource
Tweet contains Ebola news, updates, or information. May be the title or summary of the linked article. Contents may or may not be factual.
<u>Example:</u> 'Ebola outbreak the worst of its kind(news): tinyurl'

2=Personal Experience
Twitter user mentions a direct (personal) or indirect(e.g. friend, family, co-worker) experience with Ebola or the social/economic effects of Ebola.
<u>Example:</u> 'Ebola had me concerned about travelling when it first broke out but when I found out it was not airborne I was fine'

3=Personal Opinion and Interest
Twitter user posts their opinion of the Ebola outbreak or expresses a need for or discovery of information. General Ebola chatter or commentary.
<u>Example:</u> 'Ebola is not a threat to the UK I don't know why the UK media and news or reporting it so much!'

4=Jokes/Parody
Tweet contains an Ebola joke delivered via video, text, photo, or a humorous opinion of Ebola that does not refer to a personal experience
<u>Example:</u> 'Hope he gets Ebola.' 'He coughed so hard he gave us Ebola.'

5=Marketing
Tweet contains an advertisement for Ebola related product or service
<u>Example:</u> 'Buy face masks now and protect yourself from Ebola: tinyurl'

6=Spam
Tweet is not related to Ebola. Also include 'chain-tweets' (would be good to explain what you mean by chain tweets just to make it as clear as possible because some people may not understand it).
<u>Example:</u> 'Big data, spaceships in 2020 to use big data to reach light speed #ebola' 'RT if you don't want Ebola.'

7=Ambiguous
Tweet contains two categories or does not fall into any of the categories above.
Example: 'Ebola and cancer words make me throw up'

Tweets not resource or spam based were placed into qualifier categories:

2 Secondary Code Frame

Note: Only code this for tweets which were not resource or spam tweets. For resource and spam tweets fill in 0.

1=Humour/Sarcasm
Tweet is comedic or sarcastic
Example: 'The man was coughing so much I think he gave us all Ebola'

2=Relief
Tweet expresses joy, happiness, or sense of peace
Example: 'Ebola is going Relief!'

3=Downplayed Risk
Tweet attempts to de-emphasise the potential risk of Ebola or bring it into perspective. May also express a lack of concern or disinterest
Example: 'Ebola can't survive in the UK, the risk is being overplayed'

4=Concern
Tweet expresses Ebola-related fear, anxiety, worry, or sadness for self or others. May also express scepticism.
Example: 'My family is going on holiday to a high risk Ebola zone hope they don't get it! So scared'

5=Frustration
Tweet expresses anger, annoyance, scorn, or volatile contempt. May include coarse language
Example: 'So they invent a cure for Ebola like so quick when it came to the west #doublestandards'

6=Misinformation/Speculation
Tweet contradicts the reference standard or contains unsubstantiated information. May make speculations or express distrust of authority or the media. May include conspiracy or doomsday theories.
Example: 'Ebola, Bird Flu, ISIS, the end is near!'; 'The government made Ebola to make money from us all: tinyurl.com'

7=Question
Tweet asks a question or contains a question mark.
Example: 'Why did the Simpsons reference Ebola in 1997? How long has it been around?'

8=Ambiguous - Tweet does not fall into any category above or may fall into more than one category
Example: 'Shoot the Ebola patient'

9=References popular culture

Connects are makes reference to popular culture, TV, Film, or Video games
Example: ‘Ebola outbreak is like the Zombie Apocalypse’; ‘The Simpsons mentioned the Ebola outbreak’

10=Fear
Tweet promotes fear
Example: Terrorists with Ebola are coming to the U.S.

Appendix 3 Tweets outside time period for swine flu

A.1 General Fear

The search terms that were used were: 'Afraid OR Scared OR Fear AND Swine Flu' for each month of the outbreak, and the table below highlights their occurrence.

Time Period	Tweet illustration
May 2009	'I am unable to sleep always thinking about Swine Flu' [4th May]
June 2009	'I am still ill and feel it could be Swine Flu' [29 June]
July 2009	'Don't want to be thinking about Swine Flu – it scares me' [7th July]
August 2009	'Since I heard that there was Swine Flu at work I feel scared'[28 August]]
September 2009	**'**Swine Flu has been observed here! I am so scared now**'** [15 September]
October 2009	'Person next to me is coughing – I hope I don't catch Swine Flu' [October 8]
November 2009	'My child has a fever, and now scared about Swine Flu!' [November 28]

C.6 Name Discussion

The search terms that were used was 'Swine Flu AND Name'.

Time Period	Tweet illustration
May 2009	'Wonder if H1n1 needs a better name' [29 May]
June 2009	'We should name diseases after politicians rather than poor pigs!' [28 June]
July 2009	'If I was a pig I would like the name Swine Flu!' [21 July]
August 2009	'Does anyone know how H1H1 influenza got to be known as Swine Flu?' [26 Aug]
September 2009	'Swine Flu is a very bad name' [24 Sep]
October 2009	'I have a new name for Swine Flu – pig flu!' [25 Oct]
November 2009	'The name Swine Flu is disgusting – hope I don't get it' [13 Nov]

D.4 Media Organisations (critical)

The search term used in order to locate tweets which mentioned Swine Flu and which were critical of the media were: 'Swine Flu AND Media'.

Time Period	Tweet illustration
May 2009	'All this exaggeration over Swine Flu by the media is scary ' [28 May]
June 2009	'Swine Flu appears to be a big media scare' [29 June]
July 2009	'Media are hyping up the Swine Flu vaccine again' [30 July]
August 2009	'Media scare tactics continue!' [21 Aug]
September 2009	**'**Difference between regular flu and Swine Flu – is panic by media' [4 Sep]
October 2009	'I have to say I think the media has really blown this Swine Flu way out of proportion' [30 Oct]
November 2009	'Actually it might be harder to catch Swine Flu as what the media make out!' [29 Nov]

F.2 Reference to Mexicans

The search term used in order to locate tweets were: 'Swine Flu AND Mexicans'.

Time Period	Tweet illustration
May 2009	'A tiny Mexican girl is sitting next to me – she stinks! Now I know the smell of Swine Flu' [22 May]
June 2009	'The people affected by Swine Flu were young Mexicans according to this [URL]'
July 2009	''I am in love with a Mexican girl – of all countries it had to be Mexico where the Swine Flu is! ' [22 July]
August 2009	Back from a Mexican break – was sweating like a pig when over there – seems right for a place with Swine Flu [30 Aug]
September 2009	**'**Just arrived back from Mexico, I don't have Swine Flu and I was kissing so many Mexicans!**'** [25 May]
October 2009	'Saw a Mexican, he was dressed informally – open toe sandals! Then they worry why they have Swine Flu' [28 Oct]
November 2009	'Survived the Swine Flu – F*** the Mexicans!' [8 Nov]

G.1 Pork Consumption

The search terms used were 'Swine Flu AND Pork'.

Time Period	Tweet illustration
May 2009	I have not eaten Pork since Swine Flu began! Yesterday I had bacon ribs and I am still alive. [26 May]
June 2009	'I have a cough , it could be Swine Flu – I will eat some more Pork to spite those pigs' [29 June]
July 2009	'Someone I know believed that if they ate Pork they would get Swine Flu!' [29 July]
August 2009	'Even after a Swine Flu outbreak – people still are not giving up their Pork!' [Aug 3]
September 2009	**'**I bet Pork sales are dropping! Was going to get a tacos and was told not to because I'd get Swine Flu**' [**Sep 18**]**
October 2009	'Wow – with all of this Swine Flu news – I need to stop eating Pork' [28 Oct]
November 2009	'I got my Swine Flu vaccine! Now I can have Pork' [20 Nov]

H.3 Popular Culture/Understanding

The search term used was 'Swine Flu AND Zombies OR zombie apocalypse OR apocalypse'.

Time Period	Tweet illustration
May 2009	'I am definitely convinced that the Swine Flu outbreak is the start of the Zombie apocalypse' [4 May]
June 2009	'Swine Flu alert increased and now it is time for the zombie apocalypse!' [11 June]
July 2009	'Just heard about a Swine Flu death! Let us all get ready for the zombie apocalypse' [10 July]
August 2009	'When Swine Flu finally mutates – then it will be the start of the zombie apocalypse' [4 Aug]
September 2009	'Many of my colleagues have either the Zombie Apocolypse or the Swine Flu. I need to find myself a gun' [11 Sep]
October 2009	' Swine Flu outbreak is beginning to sound a lot like a zombie apocalypse' [7 Oct]

Appendix 4 Tweets outside time period for Ebola

Tweets which matched Ebola related keywords for the following search terms: 'Ebola AND Fear OR Scared OR Afraid'.

E.2 Fear

Time Period	Tweet illustration
April 2014	'Yeah that's the true, also I am scared after hearing about the Ebola outbreak'
May 2014	'Everyone who states they are not afraid of Ebola is lying!'
June 2014	'One of my largest fears is Ebola'
July 2014	'Hearing about Ebola has got me scared'
August 2014	'Anyone else scared of Ebola? I just want to cry'

H.11 Conspiracy Theories

Tweets were located by using the following search terms: 'Ebola AND us government OR conspiracy OR bioweapon OR manmade'.

Time Period	Tweet illustration
April 2014	'Ebola outbreak is because of population control by the government'
May 2014	'Government has neglected us all these years Ebola has been occurring'
June 2014	'Wonder what the government has done to make sure Ebola will not hit us'
July 2014	'The U.S. government is behind Ebola'
August 2014	'All of these lies and propaganda from our governments!'

E.1 Obama

The keywords used to retrieve tweets were based on the following search terms: Obama and Ebola.

Time Period	Tweet illustration
April 2014	'Thanks to Obama Ebola is spreading'
May 2014	'Obama is the Ebola president'
June 2014	'Obama admin will deny the scope of the Ebola outbreak'
July 2014	'Don't have faith in Obama to fight the Ebola'
August 2014	'Obama plotted for the occurrence of the Ebola outbreak'

J.3 Zombies

The keywords used to search tweets consisted of the following search terms: Ebola and Zombies.

Time Period	Tweet illustration
April 2014	'Ebola is spreading now – we will have real zombies'
May 2014	'Wow – have been checking the number of Ebola deaths. #zombies'
June 2014	'Hearing about Ebola in the UK – this is the start of the zombie apocalypse'
July 2014	'Zombies are coming due to Ebola outbreak'
August 2014	' I bet the Ebola vaccine will turn everyone into zombies '

H.5 Western Privilege

The keywords used to search tweets consisted of these search terms: Ebola and white people.

Time Period	Tweet illustration
April 2014	'White people are animals, and this Ebola was put here on purpose'
May 2014	'Think the white people manufactured Ebola'
June 2014	'Imagine how upset people would have been if it was white people who were dying because of Ebola'
July 2014	'Only 40 more white people need to die and we will have a Ebola vaccine'
August 2014	'So funny because the government did not care when all those until white people started getting Ebola '

Appendix 5 Tweets outside time period for Zika

Name Discussion

Time Period	Tweet illustration
Jan 2016	"Feel sorry for Tata – wonder why they selected the name after an African forest"
Feb 2016	"Guessing for sure that Zika is the worst baby name this year"
March 2016	"The name of Zika derives from a Forest in Uganda"
April 2016	"Zika has a adorable name"
May 2016	"I thought that Zika was a name of mascot"
June 2016	"Zika has the name of a song"
July 2016	"When you realise Zika is not named after a stripper"

Information Seeking

Time Period	Tweet illustration
Jan 2016	"How many Zika cases were reported in the Middle East?"
Feb 2016	"What is Zika?"
March 2016	"Can we say the Zika is the new Malaria?"
April 2016	"How can we help Latin America with the Zika virus?"
May 2016	"What are the effects of Zika?"
June 2016	"Can animals spread the Zika"
July 2016	"Is Zika still a problem?"

Olympics Rio 2016

Time Period	Tweet illustration
Jan 2016	"Zika seems to be spreading across the Americas – fear for Olympics"
Feb 2016	"CDC warns that Zika will be a risk at the Olympics"
March 2016	"Brazil suggests travels should not avoid Olympics because of Zika"
April 2016	"An anti-zika uniform for the Rio Olympics has been released "
May 2016	"A 100 academics request for the Olympics to move because of Zika"
June 2016	"They need to cancel the Olympics because of Zika"
July 2016	"How are they doing the Olympics in Rio! There is the Zika "