Understanding the Patterns of Health Information Dissemination on Social Media during the Zika Outbreak

Xinning Gui, MA[1], Yue Wang, MS[2], Yubo Kou, PhD[3], Tera Leigh Reynolds, MPH, MA[1], Yunan Chen, PhD[1], Qiaozhu Mei, PhD[2], Kai Zheng, PhD[1]
[1]University of California, Irvine, Irvine, CA; [2] University of Michigan, Ann Arbor, MI; [3]Purdue University, West Lafayette, IN

Abstract

Social media are important platforms for risk communication during public health crises. Effective dissemination of accurate, relevant, and up-to-date health information is important for the public to raise awareness and develop risk management strategies. This study investigates Zika virus-related information circulated on Twitter, identifying the patterns of dissemination of popular tweets and tweets from public health authorities such as the CDC. We leveraged a large corpus of Twitter data covering the entire year of 2016. We analyzed the data using quantitative and qualitative content analyses, followed by machine learning to scale the manual content analyses to the corpus. The results revealed possible discrepancies between what the general public was most interested in, or concerned about, and what public health authorities provided during the Zika outbreak. We provide implications for public health authorities to improve risk communication through better alignment with the general public's information needs during public health crises.

Introduction

Social media have become important information hubs, where individuals and organizations create and disseminate real-time content beyond their personal social networks and physical location. As a result, social media have become a powerful channel for information seeking and sharing, especially in times when timely information is critical, e.g., during crisis breakout[1,2]. An increasing number of public health authorities are seizing this opportunity to improve risk communication during public health emergencies (e.g., disease outbreaks, natural disasters) [3,4]. For instance, during the H1N1 influenza pandemic, governmental institutions such as the U.S. Centers for Disease Control and Prevention (CDC) and Department of Health & Human Services (HHS) disseminated information to the general public through social media and other websites[5,6]. Similarly, during the 2011 South East Queensland floods crisis, state and local authorities shared information and updates through this avenue[7]. The existing research assessing using social media for risk communication in practice, suggests that social media can be a powerful channel for communication during crises. For example, several studies have shown that using social media is highly beneficial for effective and fast risk communication during disease outbreaks and disasters[2,4,7,8].

Although many public health organizations use social media in risk communication, it is unclear whether the general public actually receive such information properly[3]. More evaluations of the effectiveness of crisis-specific risk communication efforts are needed for developing best practices in social media use[9]. In this sense, research that investigates the distribution and determinants of public health messages on social media has important implications. First, public health organizations can monitor and evaluate the general public's attitudinal and behavioral reactions during emergency situations. Second, public health organizations and practitioners can learn to adopt more effective communication strategies to improve the dissemination of health-related knowledge, announcements, and recommendations[2]. For example, by identifying a surge of misinformation, public health professionals can adapt their communication efforts to counterbalance the misinformation on social media[10]. Similarly, investigating surges in information demand can provide insights on common information gaps that exist and, thus, inform what kinds of information to disseminate[10].

In this study, we analyzed risk communications during a 2016 public health emergency, the Zika virus outbreak, on one social media platform, Twitter. Our goal was to understand how public health authorities, and the general public, communicated on Twitter during the Zika virus outbreak, and to provide implications for effective risk communication efforts. Most people, at least outside of infectious disease physicians and researchers, had not heard of the Zika virus prior to the recent, highly publicized outbreak. On February 1, 2016, the World Health Organization (WHO) declared a cluster of infants born with microcephaly and other neurologic disorders in Brazil to be a public health emergency of international concern (PHEIC)[11] . WHO noted that, while Zika virus infection was strongly suspected to be the cause, it had not yet been scientifically proven [11]; thus, declaring the cluster to be a PHEIC signaled the need for an urgent international effort to fill this research gap[12,13]. By November 18, 2016, WHO reported that there was now sufficient evidence that Zika virus infection was the cause, and ended this PHEIC[12]. While this likely resulted in reduced media and public attention, it does not mean that the Zika virus is no longer a public health threat or priority. On the contrary, WHO described the consequences of Zika virus infection as "a highly significant long-term problem" that requires a "robust longer-term response mechanism," and stated that WHO was shifting focus and resources accordingly[12]. Therefore, public health organizations will most likely continue to communicate risk information related to the Zika virus.

Based on existing research, Twitter is the best social media platform for our research goal. Among the various platforms, Twitter has frequently been used by public health authorities, such as WHO, the U.S. National Institutes of Health (NIH), and local health departments, to publish information about disease outbreaks[14–16]. Therefore, it is very likely that many public health organizations also disseminated risk information during the Zika virus outbreak. Recent research also shows the great potential of mining and analyzing Twitter data[2] that can be used as "a proxy measure of the effectiveness of public health messaging or public health campaigns"[17].

We collected a 10% sample of all tweets that contain "Zika" and were circulated on Twitter throughout 2016 (from January 1st to December 31st), as this was the time period when Zika obtained the most international attention (due to WHO's announcement). We analyzed the typology of these tweets and their relation to tweets that were the most retweeted. Using the same strategy, we also analyzed tweets that came from authoritative accounts. Authoritative accounts are those that are affiliated to public health organizations. Unlike in noncontagious diseases, epidemics and pandemics necessitate effective and timely communications from public health organizations, as they need the general public's cooperation to mitigate the situation. Our case study of risk communication during the Zika virus outbreak can shed light on how public health organizations could (1) improve their ongoing Zika-related communications, and (2) better leverage Twitter to target the general public's' information needs, communicate more effectively and reach more people in shorter in shorter amount of time during public health crises.

Methods

We used a mixed-methods approach that combines quantitative and qualitative content analysis methods to analyze a sample of tweets. To scale up the mixed approach to a large collection of tweets, we leveraged machine learning to assist in qualitative coding. Our general analysis flow contains seven steps (see Figure 1 for a flow chart):

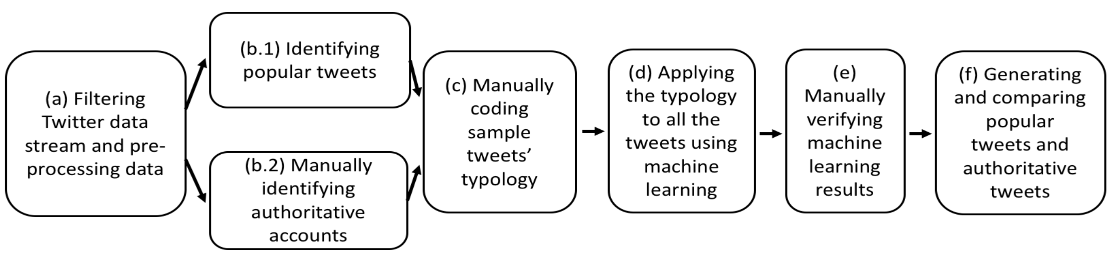

Figure 1. Flow of our methodology. Each block is described in detail in text.

(a) Through Twitter's gardenhose API, we can access 10% of the entire Twitter data stream. We prepared our dataset by selecting tweets that contained the keyword "Zika" in the year 2016. By doing this, we yielded a dataset consisting of 1,495,480 tweets. Only English tweets were used in the manual content analysis. English tweets account for 54% of all Zika-related tweets.

(b.1) We defined "popular tweets" to be those that had been retweeted at least 100 times. We identified popular tweets based on Twitter's "retweet" data field. This gave us 3,581 unique tweets, which we refer to as the "popular set."

(b.2) We sorted all the accounts ranked by"# of mentions" (using @ token), and manually selected 17 authoritative accounts that are dedicated to public health issues, and represent public health agencies, medical experts, and institutions from top 300 accounts. Though these accounts themselves do not publish many tweets, they are frequently mentioned (using @ token). Table 1 shows the top mentioned authoritative accounts. During the accounts selecting process, we excluded some indirect accounts such as POTUS (the U.S. president's Twitter account) and JuanOrlandoH (Official account of the President of the Republic of Honduras). These 17 authoritative accounts gave us 1,227 unique Zika-related tweets, which we refer to as the "authoritative set."

(c) We sampled 200 tweets from the popular set. From the authoritative set, we sampled five tweets from each of the 17 authoritative accounts. In total, we had 285 tweets. Employing a grounded-theory approach[18], each of the two authors coded the content of these 285 tweets. They first read the tweets to get a general sense of the content. After that, each author followed the open coding process to code the 285 tweets and generate a set of initial codes independently. Through several rounds of reading, coding, and comparing emerging data to existing themes, the two authors generated a set of codes that emerged naturally to describe the tweet content. The two authors discussed and consolidate the codes through the axial coding process. The final codebook contains a set of 8 codes for the sample of 285 tweets.

Table 1: Taxonomy of 17 authoritative accounts we selected.

Category	Definition	Selected Twitter Accounts
CDC official	Current CDC scientists and spokespeople	CDCgov ; DrFriedenCDC ; CDCMMWR ; CD Cemergency ; CDCtravel
WHO official	Current WHO scientists and spokespeople	WHO
Federal official	Current or former federal officials, administrators and politicians	HHSGov; SecBurwell; Surgeon_General76Y
Academic or medical expert/Institution	Healthcare practitioners, medical researchers/institution and academic scientists	MackayIM; PositiveID_Corp; NIH; greg_folkers; CIDRAP
State or local official	Current or former local and state officials, administrators and politicians	FLGovScott; MayorGimenez; HealthyFla

(d) We used a machine learning approach to generalize from the manually coded tweets and assign codes to the remaining tweets. Using coded tweets as the training set, we developed a text classification model. The model was a one-versus-rest multiclass logistic regression classifier, taking in the content of a tweet, and predicting the most likely category. The preprocessing steps included lowercasing, removing common punctuations (".", "," and ";") and normalizing URLs (replacing by the token "URL"). Each tweet was represented as bag-of-unigram feature vectors with binary values. The classifier was trained using LibLinear package, with default regularization weight $C = 1$[19].

Table 2 shows each code's definition, example, frequency in the sample, frequency in the whole dataset, and top informative content keywords with highest information gain in each category.

Table 2. Typology of tweet content related to the Zika virus (continued).

Category	Definition	Example	Coded	Predicted in all tweets	Top informative keywords
Joke	Tweets intended to create amusement or laughter	We're worried about athletes being exposed to Zika in Rio but nobody thinks about 4 years from now when they'll be exposed to anime in Japan	21	21,778	worried, trump, usa, net, indian, fans, scream, build, joked, president
Policy	Statements or questions concerning policymaking and politicians' actions.	#Zika funding isn't a place for your selfish, partisan games, @SenateGOP. People's lives are at risk. Bring a clean bill to a vote NOW.	58	145,784	bill, women, funding, clean, GOP, help, pass, #zika, virus, hurt
Research Progress	Sharing latest news about research results on Zika and related research gaps.	Tackling Zika—using bacteria as a Trojan horse - https://t.co/aXlvF1yYzU	47	100,801	scientists, world's, pledge, race, findings, disease, top, sequence, researchers, fetus
Infection Update	Sharing news about latest infection area and travel warnings.	Nine expecting Us residents confirmed as getting Zika virus... - https://t.co/5512F27ZBE https://t.co/yK9bSf2tho	45	235,009	travel, CDC, reports, case, confirmed, nine, sexually, Texas, transmitted, first
Sports Events	Tweets concerning the relationship between sports events and Zika.	Can the Rio Olympics 2016 survive Brazil & Zika virus outbreak?... - https://t.co/YtHM9nO1aK https://t.co/RiMptBhGoe	10	3171	usa, fans, scream, joked, dangerous, time, goalkeeper, she, ball, rio

Category	Definition	Example	Coded	Predicted in all tweets	Top informative keywords
Consequence	Latest findings on the severe consequence of Zika infection.	Zika Virus could confirm deadly with 3 deaths joined to suspected troubles... - https://t.co/Tspm6oT5ho https://t.co/ZP1Jr3GkO6	18	10274	link, microcephaly, cause, zikas, shows, deaths, inadequate, hijacked, testing, thwarts
Scientific Knowledge	Introducing well-established knowledge about Zika, such as disease symptoms and preventive measures.	Q: What is #Zika? A: https://t.co/3Fq6lJSQpz #ZikaVirus https://t.co/3cUlnQgloR	78	276,020	#CDCchat, #zika, infection, know, contentmine, evidence, #pregnant, confused, tip, symptoms
Pharmaceutical Progress	Latest pharmaceutical research progress and new findings such as vaccine research progress.	GeoVax $GOVX will hit #WSJ front page: as a #Zika #vaccine swoops the world before #REO2016 #Olympics to a create #trading powerhouse. #GOVX	8	1,408	vaccine, against, swoops, trials, responses, #govx, protection, offers, geovax, immune

(e) We manually checked randomly sampled tweets with those applied codes to ensure the accuracy of the prediction. To ensure that the text classification model generates accurate predictions, we also conducted 10-fold cross-validation on the 285 coded tweets to evaluate the quality of the text classification model. Cross-validation is a commonly used statistical procedure for estimating the quality of machine learning models on uncoded data instances. For multi-class classification, the accuracy, macro-averaged precision, and macro-averaged recall are defined as

$$\text{accuracy} = \frac{\sum_{y\in Y} TP_y}{N} \qquad \text{macro-averaged precision} = \frac{1}{|Y|}\sum_{y\in Y}\frac{TP_y}{P_y} \qquad \text{macro-averaged recall} = \frac{1}{|Y|}\sum_{y\in Y}\frac{TP_y}{N_y}$$

Where Y is the set of 8 categories; TP_y is the number of tweets correctly predicted to belong to category y; N is the total number of tweets; P_y is the number of tweets predicted to belong to category y; N_y is the number of tweets actually belong to category y. Our text classifier achieved 79.6% accuracy, 75.2% macro-averaged precision, and 74.0% macro-averaged recall on the multi-class classification task, which is reasonably good. These predicted codes were used as a basis for further content analysis.

(f) We generated and compared typological difference between popular tweets and authoritative tweets.

Results

Temporal and Geographical Characteristics of Zika-related Tweets

We first describe general characteristics of the whole zika related tweets. As the word "Zika" is shared across languages, these tweets are sent from regions speaking diverse languages, where English (54%), Spanish (27%), and Portuguese (12%) are among the top three. Figure 2 shows the timeline of these tweets, annotated with important events. To cross-validate our data collection, we also show the Google Trend of "Zika" for the entire year. The two timelines are consistent. We found that tweets peaked often right after a major event. The tweet amount peaked around late January, during which locally transmitted infection had been reported from more than 20 countries. The severity of the outbreak triggered a massive amount of discussions on Twitter. In July and August, Twitter users paid more attention to Zika again, as the 2016 Olympic Games took place in August in Brazil, the place where the Zika outbreak originated from. Interestingly, the trajectory of Tweet amount rose and dropped sharply during the first peak, possibly because the initial shocking news lasted very briefly. By contrast, the trajectory of Tweet amount during the Olympic Games covers a much longer period of time with gradual increase and decrease, possibly because the general public's attention to the Olympics spanned over longer time and discussion continued prior, during and after it, since the Olympic Games was a major event that lasted for weeks.

Using 1% of these tweets that have GPS information, we visualize the worldwide discussion of Zika in an evolving heatmap. Due to the scope and space of this paper, we only show two screenshots from mid-January and mid-February (see Figure 3).

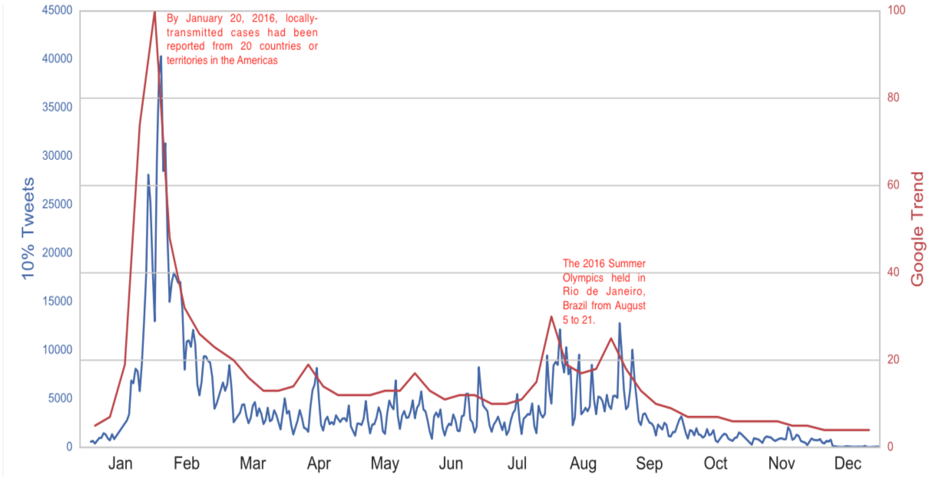

Figure 2. Timelines of Zika tweets volume and Google Trend. The left y-axis shows the number of tweets in our data set; the right y-axis shows the search volume index of Google Trend (normalized between 0 and 100).

The discussion was heated in the first three months of 2016. It originates from countries nearby the tropical region, especially South American countries. At the end of January, it has attracted attention in many countries, and swept the world in February. We notice that the spots on heatmap correlates with monthly infection update worldwide. By mid-February, Zika cases had been reported in Latin and Central America, the United States, Africa, Southeast Asia, among other countries and regions [20].

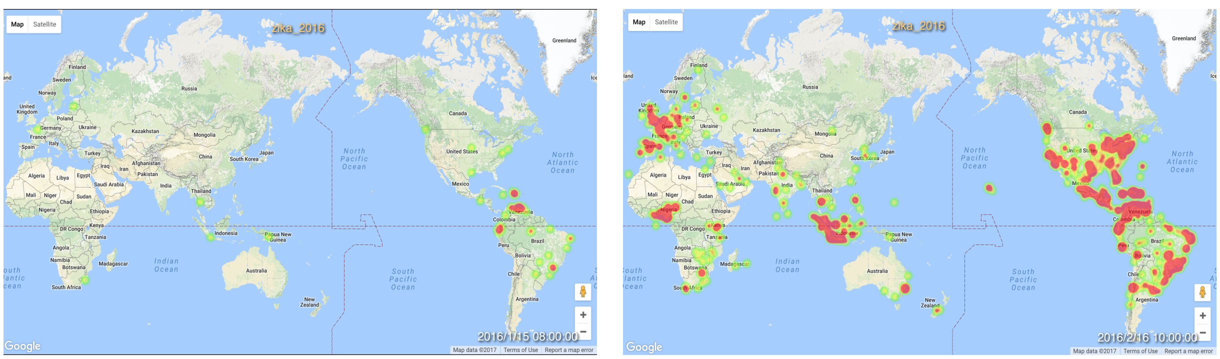

Figure 3. World heatmap of Zika-related tweets volume (mid-January and mid-February). Red regions are places where Zika-related tweets were concentrated.

Typology of Zika-related Tweet Content

We classified all the English tweets into eight categories. Within the typology, joke and sports events are the least relevant to the Zika virus itself, compared to other categories that directly point to scientific discovery and effective countermeasures. A closer read of tweets in these two categories suggested that joke was prevalent possibly because of the characteristics of Twitter as a source of entertainment; and sports events was also popular mostly because the 2016 Olympic Games were held in Brazil, where the Zika started to spread. We posit that during a health crisis, critical information such as infection update, knowledge, and progress should be disseminated as broadly as possible. Knowing the basic types of tweet content, it is crucial to understand whether critical information about health crisis is related to the tweet's popularity, in terms of retweet number and like number. Based on the typology, we explore descriptive statistics of these categories to understand each category's distinct characteristics. Table 4 shows each type's length, retweet number, and like number.

Table 4 shows that jokes were the most popular type of content on Twitter, in terms of retweet (avg=1362) and like numbers (avg=1425). We suggest that it is possibly because Twitter's one primary function is to entertain. This finding correlates with other work reporting humor as a major type of content on social media[21]. Such information might have negative consequences to Zika-related public communication, as the amusement caused by these widespread jokes might dilute the general public's awareness and caution of the severity of Zika infection. For example, the top retweeted tweet was an English translation of a Portuguese tweet:

BrazilStats2: USA goalkeeper joked about Brazil being dangerous due to Zika. Every time she hit the ball, Brazil fans scream ZIKA. https://t.co/zdZnVYqV4R

Table 4. Descriptive Statistics of Top Retweeted Tweets (# of retweets >= 100).

Category	Retweet count average (std. dev.)	Like count average (std. dev.)	Percentage (%)
Joke	1362 (3706)	1425 (3040)	1.6
Policy	289 (341)	376.8 (683)	34.2
Research Progress	419 (672)	301.4 (529)	9.5
Infection Update	228 (260)	126.6 (196)	31.9
Sports Events	336 (452)	317.0 (311)	0.4
Consequence	704 (922)	450 (680)	1.5
Scientific Knowledge	253 (381)	197 (399)	20.6
Pharmaceutical Progress	1192 (1509)	607 (787)	0.4

The Twitter account then only had 15771 followers, but the tweet was retweeted 21982 times. Noticeably, among these top retweeted tweets, factual information related to policy, infection update, and scientific knowledge had the least popularity. Each of these three received averagely less than 400 retweets and less than 570 likes. For example, an authoritative "infection update" tweet from CDCMMWR that was aimed to disseminate latest report but only retweeted twice, possibly because the reports are lengthy and require sufficient health literacy to understand:

CDCMMWR: Read @CDCgov's latest #Zika reports on the @CDCMMWR website: https://t.co/IwLglRQKol https://t.co/nsoP6vhlur

Such huge discrepancy in popularity between different types of content signifies Twitter's ineffectiveness in disseminating critical health information during a public health crisis.

After discussing popular tweets' characteristics, now we examine authoritative tweets. In Table 5 we discuss the tweets from 17 authoritative accounts. Among the 1,277 authoritative tweets, only 27 are retweeted more than 100 times ("popular"). We found that authoritative tweets do not contain content related to joke and sports events. This phenomenon correlates with our previous interpretation that joke and sports are the least relevant to Zika itself. Public health authorities might consider it inappropriate and does not match their authoritative identity.

Table 5. Descriptive Statistics of Authoritative Tweets (continued).

Category	Retweet count average (std. dev.)	Like count average (std. dev.)	Percentage (%)
Joke	0 (0)	0 (0)	0
Policy	26 (38)	19 (30)	17.0
Research Progress	17 (32)	12 (24)	9.1
Infection Update	25 (45)	87 (56)	24.8
Sports Events	0 (0)	0 (0)	0
Consequence	34 (73)	14 (23)	0.7
Scientific Knowledge	43 (177)	22 (65)	47.6

Category	Retweet count average (std. dev.)	Like count average (std. dev.)	Percentage (%)
Pharmaceutical Progress	73 (94)	34 (44)	0.2

Comparing Table 4 and Table 5, we noticed categorical differences between popular tweets and authoritative tweets. People retweet jokes and sports events, which can be problematic in conveying the critical information about health crisis, especially when those jokes and events include unverified information. Other than jokes and sports events, the general public was interested in research progress, consequence, and pharmaceutical progress, in terms of numbers of retweet and like numbers. However, most of the authoritative tweets fall into the categories of policy, infection update, and scientific knowledge. This difference suggests that in terms of public communication on Twitter, it seems that the general public's attention and the information that authoritative accounts attempt to disseminate diverge significantly: the general public is seeking and broadcasting information that tells them the results of Zika infection and what can be done against it. However, public health authorities seem to tend to publish objective, verified information in a scientific, rational tone. This discrepancy and its influence are worthy of future research.

Temporal Development between Popular and Authoritative Tweets

We compare the difference between popular and authoritative tweets in terms of their typological distribution. Figure 4 shows each type of tweets' development across time, where popular tweets are highlighted in blue and authoritative tweets in red.

As shown in Figure 4, popular tweets mostly correlate with the overall trajectory in Figure 2, peaking from January to early February and in early August. This is understandable because social media oftentimes mirror the general public's attitude, concern, and interest in real time. However, authoritative tweets do not correlate with the temporal trajectory of major Zika events (See Figure 2.). Public health authorities broadcasted information at a regular pace that did not match the general public's interest and concern. This can be easily observed in the diagrams of policy, research progress, infection update, scientific knowledge, and pharmaceutical progress. Public health authorities also tweeted more about consequence at the beginning of 2016, possibly because such information was still new to the general public at that time. However, they stopped broadcasting such information later. This suggests that when public health authorities used Twitter, it seems that they showed little timely sensitivity to major news and events. It is also possibly because public health authorities possess limited capacity in generating popular, or "viral," content.

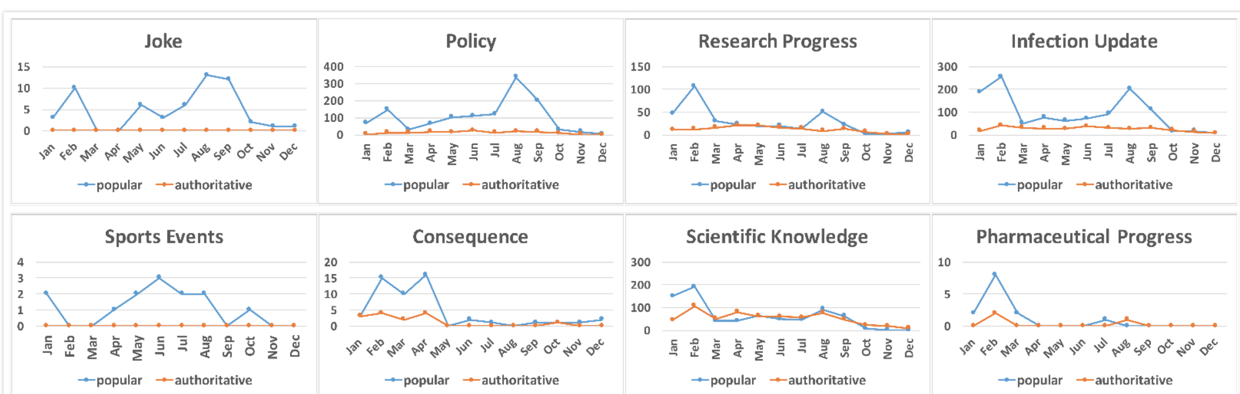

Figure 4. Temporal development of popular and authoritative tweets. All y-axes show the number of tweets in a specific category as predicted by the machine learning classifier.

Discussion

Our results indicate that Twitter does support widely and real-time communication during public health emergencies. It was widely used by the general public to discuss the Zika virus outbreak on a timely manner, suggesting that it is an important channel during public health emergencies.

By analyzing Zika-related tweets, we found that both the general public and public health authorities actively used Twitter as a platform to circulate Zika-related information. However, authoritative accounts constantly published factual information such as infection update and established scientific knowledge, and their tweets rarely gained as much popularity as tweets published by the general public, suggesting there may be discrepancies between what the general public is most interested in, or concerned about, and what public health authorities are providing during the risk communication processes. Indeed, the two had vastly different interests in terms of content, with the former being more interested in joke, sports, research and pharmaceutical progress, and consequence, while the latter focused on infection update, policy, and scientific knowledge. One possible reason for these differences is that, contrary to the general public, public health authorities have a responsibility to provide the most accurate information possible, are accountable for the content of their tweets and, often, must obtain approval (from multiple people) before content is released to the general public. Any error could cause unexpected harm, and erode public confidence in the institution. While these concerns are very real, our results suggest that Twitter messages from public health authorities are not reaching as broad of an audience as Twitter affords. Public agencies such as CDC use multiple ways (e.g., dedicated websites and press releases) to broadcast authoritative information, the traffic of which our collection of Twitter data cannot capture. However, the limited social media interaction and communication between the general public and public health agencies in our data might have negative consequences considering the popularity of Twitter, such as general Twitter users spreading rumors about public health situations, or even referencing pseudoscience. Previous studies of social media communication found that authorities often do not correct rumors quickly enough—at that point, they have already been widely transmitted[22–25]. During our manual coding process, we indeed identified a set of unverified information from the popular tweets published by the general public. This suggests that while social media is often imagined as a singular space where information flows freely, divisions exists, and some types of information (e.g., infection updates) only circulates in specific populations (e.g., authoritative accounts' followers), which can be a barrier to effective risk communication during health crises.

As people increasingly seek health information on social media, previous research has demonstrated the benefits and the necessity of directing people to trustworthy information sources[26–28]. Therefore, it is important to promote the accounts of public health authorities on social media, so that accurate information triumphs over unreliable information, i.e., unverified and speculative. By analyzing the dissemination of Zika-related health information, we identified implications that may improve the reach and effectiveness of public health authorities' future Zika-related communications, and public health emergencies more broadly.

First, our results show that popular tweets are often not directly relevant to Zika itself and contains noise such as jokes. This finding indicates that social media users may prefer entertaining and engaging content. Thus, rather than mainly doing one-way information dissemination, public health authorities may conduct more interactive communication strategies, such as building conversations with the general public, answering questions from ordinary citizens, and produce more engaging messages. For example, even with consideration of previously identified challenges, it is possible for public health authorities to produce more engaging content, as evidence by CDC's viral Zombie Apocalypse campaign which teaches about emergency preparedness[29]. Also, another possible reason that public health authority messages did not gain popularity is that the messages (e.g., scientific reports) may be lengthy and requires sufficient health literacy to understand. When publishing messages on social media, authoritative accounts may consider restating scientific messages in plain language. Furthermore, our study shows that, the general public cared more about Zika's research progress, consequence, and pharmaceutical progress, probably because these are pragmatically related to their everyday lives, countermeasures, and future plans. Similarly, a study of travel decision making during the Zika crisis shows that people actively sought information from both authorities, local residents, and previous travelers in order to make informed travel decisions[30]. To better attend to the general public's information needs, authoritative accounts can engage in such conversations to both interact with the general public and correct rumors, providing more details about the research and pharmaceutical progress.

Second, based on our analysis of temporal patterns of zika-related tweets, people's interests in Zika correlated with the latest news and major events. This suggests that public health authorities may consider publishing more timely content related to influential news in order to ease the panic, and to proactively provide information regarding the general public's foreseeable concerns, such as major international sports events. Additionally, noting that Zika attracted the largest number of tweets from January to February, 2016, possibly because of its newness, we suggest that public health authorities should pay special attention to these moments as opportunities to engage and educate the public. Moreover, we found that public health authorities focused on information that had been verified or clinically proven, such as policy, infection update, and scientific knowledge. However, such information often takes time to be developed after an event. Thus, we suggest that public health authorities publish previous verified or just other related information right after influential news was published or the major event happened, which may draw more public interests and also likely to get them more engaged with the information.

Third, our analysis suggests that monitoring information dissemination trends on social media can be an effective way of understanding the general public's interests and concerns. Public health authorities might consider improving capacity in social media monitoring, gaining familiarity with major conversations and debates that take place among the general public. In this way, public health authorities may have better chances in getting involved in these public discussions and broadcasting necessary knowledge.

Fourth, our study also indicates a disconnection between the public's understanding of a public health situation and available, scientific information sources provided by public health agencies in various ways. The dissemination of rumors highlights the disconnection. It is an urgent question as to how public health authorities can raise the public's health literacy and eHealth literacy in taking full advantage of authoritative, scientific information to understand public health issues and make informed health-related decisions.

Limitation

We based our analysis and drew conclusions on a portion of data from one social media platform. Therefore, our findings and suggestions may not generalize to other social media, websites, or traditional media, or other crisis situations. Our study is formative rather than definitive. Further hypothesis tests are necessary to examine our findings.

Conclusion

This research concerns risk communication during a public health crisis induced by the Zika virus. We employed mixed-method approach and machine learning to understand the content of both popular tweets and authoritative tweets. We discussed the information gap between the general public and public health authorities and practitioners. We suggested implications that may help improve public health authorities' risk communication strategies, including providing more engaging and straightforward health message contents that attend to people's information needs, adopting more interactive communication strategies, delivering messages timely after related news and major events, conducting social media surveillance and raising the public's health literacy and eHealth literacy.

References

1. U.S. Department of Health and Human Services; Centers for Disease Control and Prevention. Crisis and Emergency Risk Communication. 2014. 424 p.
2. Chew C, Eysenbach G. Pandemics in the age of Twitter: content analysis of Tweets during the 2009 H1N1 outbreak. Sampson M, editor. PLoS One. 2010;5(11):e14118.
3. Thomas TL, Schrock C, Friedman DB. Providing Health Consumers with Emergency Information: A Systematic Review of Research Examining Social Media Use During Public Crises. J Consum Health Internet.2016;20(1–2):19–40.
4. Chou WS, Hunt YM, Beckjord EB, Moser RP, Hesse BW. Social media use in the United States: implications for health communication. J Med Internet Res. 2009;11(4):e48.
5. Ding H, Zhang J. Social Media and Participatory Risk Communication during the H1N1 Flu Epidemic: A Comparative Study of the United States and China. China Media Res. 2010;6(4):80–91.
6. Reynolds BJ. Building trust through social media. CDC's experience during the H1N1 influenza response. Mark Health Serv. 2010;30(2):18-21.
7. Bruns A, Burgess J, Crawford K, Shaw F. #qldfloods and @QPSMedia: Crisis Communication on Twitter in the 2011 South East Queensland Floods. 2012.
8. Moorhead SA, Hazlett DE, Harrison L, Carroll JK, Irwin A, Hoving C. A new dimension of health care: systematic review of the uses, benefits, and limitations of social media for health communication. J Med Internet Res. 2013;15(4):e85.
9. Glik DC. Risk Communication for Public Health Emergencies. Annu Rev Public Health. 2007;28(1):33-54
10. Eysenbach G. Infodemiology and Infoveillance: Framework for an Emerging Set of Public Health Informatics Methods to Analyze Search, Communication and Publication Behavior on the Internet. J Med Internet Res. 2009;11(1):e11.
11. World Health Organization. Zika situation report (5 February 2016). World Health Organization. World Health Organization; 2016. Available from: http://www.who.int/emergencies/zika-virus/situation-report/5-february-2016/en/

12. World Health Organization. Fifth meeting of the Emergency Committee under the International Health Regulations (2005) regarding microcephaly, other neurological disorders and Zika virus. World Health Organization. World Health Organization; 2016. Available from: http://www.who.int/mediacentre/news/statements/2016/zika-fifth-ec/en/
13. World Health Organization. Zika virus and complications. World Health Organization; 2016. Available from: http://www.who.int/emergencies/zika-virus/en/
14. Park H, Reber BH, Chon M-G. Tweeting as Health Communication: Health Organizations' Use of Twitter for Health Promotion and Public Engagement. J Health Commun. 2016;21(2):188-198.
15. Neiger BL, Thackeray R, Burton SH, Thackeray CR, Reese JH. Use of Twitter Among Local Health Departments: An Analysis of Information Sharing, Engagement, and Action. J Med Internet Res. 2013;15(8):e177.
16. Lee JY, Sundar SS. To Tweet or to Retweet? That Is the Question for Health Professionals on Twitter. Health Commun. 2013;28(5):509-524.
17. Signorini A, Segre AM, Polgreen PM. The Use of Twitter to Track Levels of Disease Activity and Public Concern in the U.S. during the Influenza A H1N1 Pandemic. Galvani AP, editor. PLoS One. 2011;6(5):e19467.
18. Corbin JM, Strauss AL. Basics of qualitative research : techniques and procedures for developing grounded theory. 4th. SAGE Publications, Inc; 2014. 456 p.
19. Machine Learning Group at National Taiwan University. LIBLINEAR -- A Library for Large Linear Classification. 2015. Available from: https://www.csie.ntu.edu.tw/~cjlin/liblinear/
20. World Health Organization. Zika situation report (19 February 2016). World Health Organization. World Health Organization; 2016. Available from: http://www.who.int/emergencies/zika-virus/situation-report/19-february-2016/en/
21. Holton AE, Lewis SC. Journalist, social media, and the use of humor on twitter. Electron J Commun. 2011 2011;21(1):1-21.
22. Andrews CA, Fichet ES, Ding Y, Spiro ES, Starbird K. Keeping Up with the Tweet-dashians: The Impact of "Official" Accounts on Online Rumoring. In: Proceedings of the 2016 ACM Conference on Computer-Supported Cooperative Work & Social Computing - CSCW '16. New York, New York, USA: ACM Press; 2016. p. 451–64.
23. Liao Q, Shi L. She gets a sports car from our donation: rumor transmission in a Chinese microblogging community. In: Proceedings of the 16th ACM Conference on Computer-Supported Cooperative Work & Social Computing - CSCW '13. New York, New York, USA: ACM Press; 2013. p. 587–97.
24. Tirkkonen P, Luoma-aho V. Online authority communication during an epidemic: A Finnish example. Public Relat Rev. 2011;37(2):172-174.
25. Kim S, Liu BF. Are All Crises Opportunities? A Comparison of How Corporate and Government Organizations Responded to the 2009 Flu Pandemic. J Public Relations Res. 2012;24(1):69-85.
26. de Boer MJ, Versteegen GJ, van Wijhe M. Patients' use of the Internet for pain-related medical information. Patient Educ Couns. 2007;68(1):86-97.
27. Syed-Abdul S, Fernandez-Luque L, Jian W-S, Li Y-C, Crain S, Hsu M-H, et al. Misleading health-related information promoted through video-based social media: anorexia on YouTube. J Med Internet Res. 2013;15(2):e30.
28. Lee K, Hoti K, Hughes JD, Emmerton LM, Platt T. Interventions to Assist Health Consumers to Find Reliable Online Health Information: A Comprehensive Review. Manchikanti L, editor. PLoS One. 2014;9(4):e94186.
29. Centers for Disease Control & Prevention (CDC). Zombie Preparedness. 2016. Available from: https://www.cdc.gov/phpr/zombies.htm
30. Gui X, Kou Y, Pine KH, Chen Y, Managing Uncertainty: Using Social Media for Risk Assessment during a Public Health Crisis. In: Proceedings of the 2016 ACM CHI Conference on Human Factor in Computing Systems - CHI '17. New York, New York, USA: ACM Press; 2017. p. 4520–4533.