Brief Report

How people react to Zika virus outbreaks on Twitter? A computational content analysis

King-Wa Fu PhD [a], Hai Liang PhD [a], Nitin Saroha BS [b], Zion Tsz Ho Tse PhD [c,1], Patrick Ip MBBS, MPH [d,1], Isaac Chun-Hai Fung PhD [e,1,*]

[a] Journalism and Media Studies Centre, University of Hong Kong, Hong Kong
[b] Department of Computer Science, University of Georgia, Athens, GA
[c] College of Engineering, University of Georgia, Athens, GA
[d] Department of Paediatrics and Adolescent Medicine, Li Ka Shing Faculty of Medicine, University of Hong Kong, Hong Kong
[e] Department of Epidemiology, Jiann-Ping Hsu College of Public Health, Georgia Southern University, Statesboro, GA

Key Words:
Health communication
public response
social media
Zika virus disease
pregnancy complications

Zika-related Twitter incidence peaked after the World Health Organization declared an emergency. Five themes were identified from Zika-related Twitter content: (1) societal impact of the outbreak; (2) government, public and private sector, and general public responses to the outbreak; (3) pregnancy and microcephaly: negative health consequences related to pregnant women and babies; (4) transmission routes; and (5) case reports. User-generated contents sites were preferred direct information channels rather than those of the government authorities.

© 2016 Association for Professionals in Infection Control and Epidemiology, Inc. Published by Elsevier Inc. All rights reserved.

The 2015-2016 Zika virus (ZIKV) epidemic in the Americas is unprecedented. ZIKV-infected pregnancy could be complicated with fetal microcephaly and long-term developmental disability.[1,2] Epidemiologic evidences suggested that ZIKV might cause Guillain-Barré syndrome.[3] The World Health Organization (WHO) declared it a Public Health Emergency of International Concern (PHEIC) on February 1, 2016.[4]

As of 2015, Twitter had 320 million monthly active users, supporting >35 languages.[5] English, Spanish, and Portuguese are among the top 10 languages used on Twitter.[6]

Our study aims to provide baseline data for future ZIKV-related Twitter health communication. We report the incidence trends of ZIKV-related Twitter data and our content analysis of a cross-sectional sample of ZIKV-related English Tweets.

METHODS

Using streaming Application Programming Interface (API), we collected a 1% Twitter sample between May 1, 2015, and April 2, 2016 (1,076,477,185 Tweets), from which we retrieved all those with the key word "Zika" (English: n = 26,439; Spanish: n = 28,342; Portuguese, n = 13,562). We then calculated the weekly (Fig 1) and daily (Supplementary Fig S1) incidence rate of Zika-related Tweets by language.

We also collected separate Twitter samples through the search API by the keyword "Zika" that reported valid language values (n = 155,341, January 26-28, 2016; 310,943, February 1-3, 2016, including re-Tweets) (Supplementary Table S1). Three data fields were analyzed: created time, body of message, and language used. Among those English, Spanish, and Portuguese Tweets carrying a Uniform Resource Locator (URL) (n = 47,242 for January 26-28, 2016; 84,791 for February 1-3, 2016), their URL's second-level domains were extracted, indicating a Tweet's direct information channel via the link.

We computationally analyzed the contents of 62,547 English Tweets obtained by search API. Topic modeling was used to group bags of words in Tweets into different topics. The 20-topic model was found to fit the data best. The first author reviewed the 20 topics
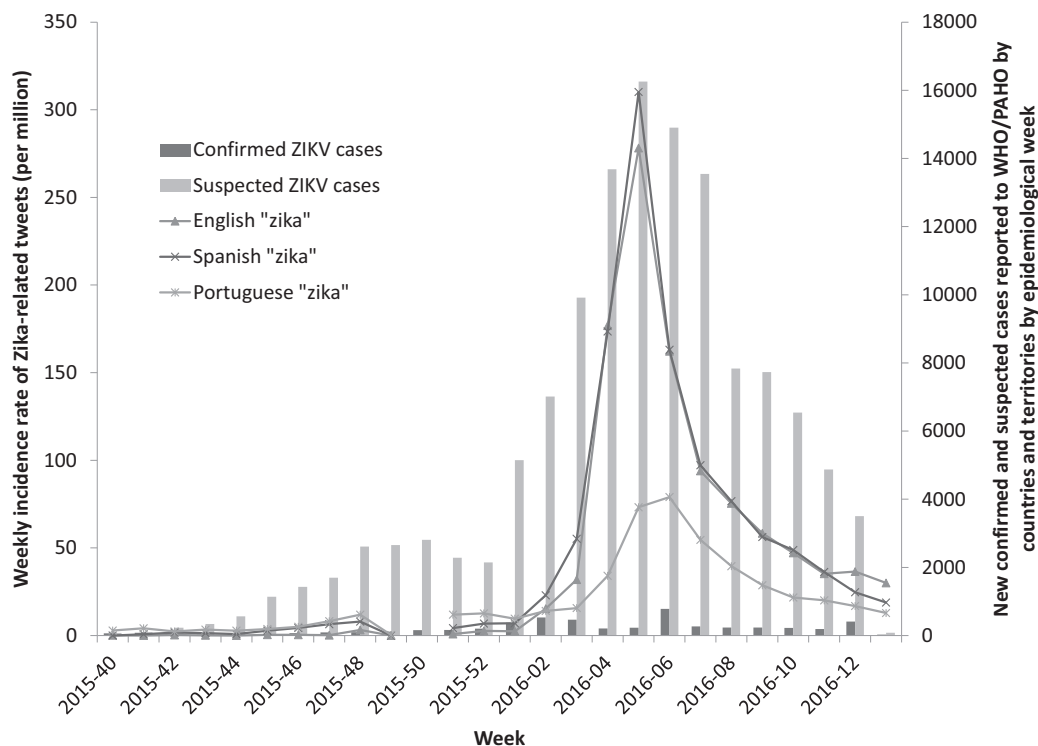
Fig 1. Weekly incidence rate of Zika-related Tweets (per million) by language (English, Spanish, and Portuguese) (left *y*-axis) and the number of new confirmed and suspected cases reported to the WHO/PAHO by countries and territories (right *y*-axis) by epidemiologic week, October 2015-March 2016. Our Twitter data were based on the 1% streaming Application Programming Interface data from May 1, 2015-April 2, 2016. Data were missing on the following dates because of server breakdown: May 20-28, 2015; June 15-18, 2015; July 2-5, 2015; July 31-August 1, 2015; August 29-September 3, 2015; November 8, 2015; November 24, 2015; December 1-7, 2015; December 9-19, 2015; December 22-23, 2015; December 30, 2015-January 4, 2016; and January 22, 2016. Weekly ZIKV-related Twitter incidence rates for weeks 48 and 49 (2015) were based on data from 2 days (November 29-30) and 1 day (December 8, 2015) respectively. Because of the missing data, cautious interpretation is warranted. Epidemiologic data were manually extracted from PAHO/WHO (http://ais.paho.org/phip/viz/ed_zika_epicurve.asp). *PAHO*, Pan American Health Organization; *WHO*, World Health Organization; *ZIKV*, Zika virus.

Table 1
Distribution of ZIKV-related Twitter content under 5 themes (n = 62,547)*

Topics	%	Representative tweets
ZIKV's impact	39.5	"3-4 million cases of Zika infection possible in the Americas over next year"
		"Olympic organizers reveal plans to combat Zika threat"
		"Zika Virus Could Become Explosive Global Pandemic"
Reaction to ZIKV	23.7	"UN to decide if Zika virus is a global health emergency"
		"United to refund travel to regions hit with Zika virus"
		"Brazil to use army against Zika virus"
Pregnancy and microcephaly	18.1	"Zika virus may be linked to a range of birth defects, not just microcephaly"
		"U.S. boosts study of Zika, birth defect link, as virus seen spreading"
		"El Salvador advises women to avoid pregnancy for 2 years due to Zika virus outbreak"
Transmission routes	10.7	"What is #Zika? Zika is a virus spread to humans by Aedes mosquitoes, the same mosquitoes that spread dengue, chikungunya, yellow fever"
		"Two Cases Suggest Zika Virus Could Be Spread Through Sex"
		"Genetically modified mosquitoes used to fight dengue, zika in Brazil"
Case reports	8.1	"Virginia resident who traveled abroad tests positive for Zika virus"
		"Zika Virus Found In Danish Tourist Returning From South America Hospital"
		"Puerto Rico Zika report to show 18 cases: health secretary"
Total	100.0	

UN, United Nations; *ZIKV*, Zika virus.
*This was an analysis of a Twitter sample with the key word "Zika" after removing embedded Uniform Resource Locator links and @ mentions (January 26-28, 2016).

and grouped them into 5 themes. The corresponding author confirmed the grouping. Representative Tweets in each topic were identified (see Supplementary Materials for details [http://healthdata.engr.uga.edu/static/publications/Zika_Twitter_2016-04-18_v8_Supp_1.pdf]).

RESULTS

Figure 1 shows the weekly incidence rate of ZIKV-related Tweets and the number of new confirmed and suspected ZIKV cases as re-

ported to the WHO. In 2015, ZIKV-related Tweets were predominantly in Spanish and Portuguese. Starting in 2016, there was a substantial increase in ZIKV-related Tweets in English and Spanish; the weekly incidence rates reached a peak of 310 (English) and 278 (Spanish) Tweets per million in week 5. The Portuguese ZIKV-related Tweets reached a peak in week 6 (79 Tweets per million).

Supplementary Table S2 presents the distribution of Tweets' URL second-level domains. The top 8 domains were all user-generated content sites, including social media platforms, such as Facebook, Instagram, Twitter, YouTube, LinkedIn, Tumblr, the blogging site

WordPress, and the question and answer site About.com. Between January 26 and 28, these 8 domains accounted for 26% of all domains. Each of the remaining 15,134 domains made up <1% of the total, representing a long-tail pattern of information platforms. The domain for the Centers for Disease Control and Prevention (CDC) (rank 90th) and the WHO Web site (rank 140th) amounted 0.06% and 0.05%, respectively. The top 2 English language news sites were those of the British Broadcasting Corporation (rank 19th, 0.24%) and Cable News Network (rank 24th, 0.18%). No apparent difference between the January 26-28 and February 1-3 samples was observed.

Our contents analysis revealed 5 themes among ZIKV-related Tweets (Table 1): (1) societal impact of the outbreak; (2) government, public and private sector, and general public responses to the outbreak; (3) pregnancy and microcephaly: negative health consequences related to pregnant women and babies; (4) transmission routes; and (5) case reports.

DISCUSSION

We reported the incidence trends of ZIKV-related Twitter traffic from May 2015-March 2016 and a computational content analysis of ZIKV-related English Tweets in late January 2016. The predominance of Spanish and Portuguese among ZIKV-related Tweets in 2015 reflected ZIKV awareness among Latin American Twitter users.[7] The rise of ZIKV-related Tweets in English in early 2016 coincided with the concerns raised in the English-language media given the WHO's PHEIC announcement, the travel-related cases in the continental United States, and the CDC's travel advice pertinent to ZIKV-affected areas.

Multilingual public health communication on Twitter is important. The CDC engages Spanish-speaking Twitter users via its Spanish profile (@CDCespanol). Likewise, multilingual travel advice will be vital as Brazil prepares for the 2016 Olympics.[8]

Our domain analysis of embedded URL links suggested that user-generated content remains the preferred direct channels of ZIKV-related information for Twitter users worldwide. Because most Twitter users would not share Web links of public health agencies' Web sites directly, social media communication may allow original official health information and health news (mostly covering official health sources) to be re-Tweeted and shared online.

ZIKV-related topics identified on Twitter resonated well with the public's concerns, given the risks of microcephaly and Guillain-Barré syndrome.[2,3] The public was concerned about the societal impact as ZIKV spread. People wanted to learn about ZIKV transmission routes and the risks associated with ZIKV infections.

Our study has its limitations. Our sample for computational content analysis was limited to English Tweets over 3 days.

Although 20 topics were identified using statistical methods, they relied on human judgement to connote them into 5 themes for interpretations. However, using both a statistical algorithm and human curators makes our study relevant to public health. Computational methods assist—not replace—health communicators during emergency responses. Our study highlights the needs of multilingual Twitter health communication on ZIKV. To understand the themes in further detail, in-depth qualitative analysis on a small proportion of Twitter content and the analysis of Spanish and Portuguese Tweets are warranted for future study.

To conclude, Twitter users shared ZIKV-related information pertinent to the epidemic's societal impact and global threat, people's responses, transmission routes, risk of microcephaly, and case reports. ZIKV-related Tweets were predominantly in Spanish and Portuguese in 2015, but English Tweets increased since early 2016, given media coverage of the PHEIC announcement and CDC travel advice. Reaching the audience with messages of ZIKA prevention and control is crucial in our combat against ZIKV.

APPENDIX: SUPPLEMENTARY DATA

Supplementary data related to this article can be found doi:10.1016/j.ajic.2016.04.253.

References

1. World Health Organization. Zika virus fact sheet. 2016. Available from: http://www.who.int/mediacentre/factsheets/zika/en/. Accessed August 9, 2016.
2. Oliveira Melo AS, Malinger G, Ximenes R, Szejnfeld PO, Alves Sampaio S, Bispo de Filippis AM. Zika virus intrauterine infection causes fetal brain abnormality and microcephaly: tip of the iceberg? Ultrasound Obstet Gynecol 2016;47:6-7.
3. Cao-Lormeau V-M, Blake A, Mons S, Lastère S, Roche C, Vanhomwegen J, et al. Guillain-Barré Syndrome outbreak associated with Zika virus infection in French Polynesia: a case-control study. Lancet 2016;387:1531-9.
4. World Health Organization. WHO statement on the first meeting of the International Health Regulations (2005) Emergency Committee on Zika virus and observed increase in neurological disorders and neonatal malformations. 2016. Available from: http://www.who.int/mediacentre/news/statements/2016/1st-emergency-committee-zika/en/. Accessed February 13, 2016.
5. Twitter Inc. Twitter usage/company facts. 2016. Available from: https://about.twitter.com/company/. Accessed August 9, 2016.
6. Mocanu D, Baronchelli A, Perra N, Gonçalves B, Zhang Q, Vespignani A. The Twitter of babel: mapping world languages through microblogging platforms. PLoS ONE 2013;8:e61981.
7. Samarasekera U, Triunfol M. Concern over Zika virus grips the world. Lancet 2016;387:521-4.
8. Elachola H, Gozzer E, Zhuo J, Memish ZA. A crucial time for public health preparedness: Zika virus and the 2016 Olympics, Umrah, and Hajj. Lancet 2016;387:630-2.