Infection, Disease & Health (2018) 23, 10—16

Available online at www.sciencedirect.com

ScienceDirect

journal homepage: http://www.journals.elsevier.com/infection-disease-and-health/

Research paper

Twitter and Middle East respiratory syndrome, South Korea, 2015: A multi-lingual study

Isaac Chun-Hai Fung [a,*,1], Jing Zeng [b,1], Chung-Hong Chan [c], Hai Liang [d], Jingjing Yin [e], Zhaochong Liu [f], Zion Tsz Ho Tse [g,2], King-Wa Fu [c,2]

[a] *Department of Epidemiology and Environmental Health Sciences, Jiann-Ping Hsu College of Public Health, Georgia Southern University, Statesboro, GA 30460-8015, USA*
[b] *Digital Media Research Centre, Queensland University of Technology, Brisbane, QLD, Australia*
[c] *Journalism and Media Studies Centre, The University of Hong Kong, Hong Kong*
[d] *School of Journalism and Communication, Chinese University of Hong Kong, Hong Kong*
[e] *Department of Biostatistics, Jiann-Ping Hsu College of Public Health, Georgia Southern University, Statesboro, GA 30460-8015, USA*
[f] *Department of Computer Science, The University of Georgia, Athens, GA 30602, USA*
[g] *College of Engineering, The University of Georgia, Athens, GA 30602, USA*

Received 26 June 2017; received in revised form 24 August 2017; accepted 24 August 2017
Available online 18 September 2017

KEYWORDS
MERS;
Social media;
Language;
Culture;
Health communication

Abstract *Background:* Different linguo-cultural communities might react to an outbreak differently. The 2015 South Korean MERS outbreak presented an opportunity for us to compare tweets responding to the same outbreak in different languages.
Methods: We obtained a 1% sample through Twitter streaming application programming interface from June 1 to 30, 2015. We identified MERS-related tweets with keywords such as 'MERS' and its translation in five different languages. We translated non-English tweets into English for statistical comparison.
Results: We retrieved MERS-related Twitter data in five languages: Korean ($N = 21,823$), English ($N = 4024$), Thai ($N = 2084$), Japanese ($N = 1334$) and Indonesian ($N = 1256$). Categories of randomly selected user profiles ($p < 0.001$) and the top 30 sources of retweets ($p < 0.001$) differed between the five language corpora. Among the randomly selected user profiles, K-pop fans ranged from 4% in the Korean corpus to 70% in the Thai corpus; media ranged from 0% (Thai) to 14% (Indonesian); political advocates ranged from 0% (Thai) to 19% (Japanese); medical professionals ranged from 0% (Thai) to 7% (English). Among the top 30 sources of retweets for each corpus (150 in total), 70 (46.7%) were media; 29 (19.3%) were

* Corresponding author. Jiann-Ping Hsu College of Public Health, Georgia Southern University, PO Box 8015, Statesboro, GA 30460-8015, USA. Fax: +1 912 478 0171.
E-mail addresses: cfung@georgiasouthern.edu (I.C.-H. Fung), zengjing310@gmail.com (J. Zeng), u3003078@connect.hku.hk (C.-H. Chan), hailiang@cuhk.edu.hk (H. Liang), jyin@georgiasouthern.edu (J. Yin), zhaochongliu@gmail.com (Z. Liu), ziontse@uga.edu (Z.T.H. Tse), kwfu@hku.hk (K.-W. Fu).
[1] These two individuals serve as co-first authors.
[2] These two individuals serve as co-senior authors.

http://dx.doi.org/10.1016/j.idh.2017.08.005
2468-0451/© 2017 Australasian College for Infection Prevention and Control. Published by Elsevier B.V. All rights reserved.

K-pop fans; 7 (4.7%) were political; 9 (6%) were medical; and 35 (23.3%) were categorized as 'Others'. We performed chi-square feature selection and identified the top 20 keywords that were most unique to each corpus.
Conclusion: Different linguo-cultural communities exist on Twitter and they might react to the same outbreak differently. Understanding audiences' unique Twitter cultures will allow public health agencies to develop appropriate Twitter health communication strategies.
© 2017 Australasian College for Infection Prevention and Control. Published by Elsevier B.V. All rights reserved.

Highlights

- Tweets about the 2015 South Korean MERS outbreak in 5 languages were compared.
- User profiles and keywords used were different across the 5 data sets.
- Health communication on Twitter should be culture and language specific.

Introduction

Different linguo-cultural communities may react to an infectious disease outbreak differently, either due to varying geographical distances from the disease outbreak [1], or cultural differences that differentiate one community's perception of the outbreak from another [2]. Outbreaks may affect populations whose primary language is not English.

Twitter corpora in different languages provide us with an opportunity to analyze the perception of and reaction to disease outbreaks among different populations. Twitter is a social media platform that allows users to communicate in different languages. According to Statistica, there were 328 million monthly active Twitter users worldwide in 2017 [3]; two-thirds of Twitter accounts were outside the US [4]; and Twitter supported 35 or more languages [5]. The transnational nature of Twitter as a social media platform where users of different linguo-cultural backgrounds can react to the same event online in real time presents us with an opportunity in which we can examine some of the hidden assumptions of social media public health literature. Many analyses of Twitter data in public health literature only studied tweets written in English. For example, a recent systematic review on social media studies about Ebola found that all seven studies on Twitter data about Ebola covered English tweets only [6]. In a recent paper on Twitter data about Zika, while the authors studied the trends of Zika-related tweets in English, Spanish and Portuguese, respectively, they limited their content analysis to English language tweets given their own language limitation [7]. However, health-related English language Twitter content may not be generalizable to Twitter content in other languages. The assumed 'global' nature of Twitter and the analysis of a monolingual corpus of Twitter data mask the underlying diverse nature of the different linguo-cultural groups of users who communicate on Twitter using different languages [8]. We contend that it is essentially an Anglophone-centric paradigm with a hidden language bias towards English [9].

The 2015 Middle East respiratory syndrome (MERS) outbreak in South Korea provided us with a unique opportunity to perform a cross-cultural comparison between different linguo-cultural communities of Twitter users [10]. Because Twitter users in the various Asian countries adjacent to South Korea write in different languages, by comparing Twitter corpora in different languages that are associated with a single infectious disease outbreak, we can control for time and event, and identify the differences between the corpora.

In this study, we are going to examine the following hypotheses.

1. The types of Twitter profiles of those who tweeted about MERS differed between the corpora of five different languages.
2. Except for keywords that are apparently specific to this outbreak (i.e., 'MERS', 'Korea', 'Korean', and 'South Korea'), there are unique keywords (when translated into English) used in MERS-related tweets that distinguish the corpora of five different languages.
3. The types of the top 30 Twitter profiles that received the most retweets for their MERS-related tweets differed between the corpora of five different languages.

By examining the aforementioned hypotheses, we will test if Twitter usage pertaining to the 2015 South Korean MERS outbreak differed along linguo-cultural lines and explore the implications of our findings for global health communications.

Methods

This is a cross-sectional study. We obtained a 1% random sample of Twitter data via Twitter streaming application programming interface (API) [11]. APIs are a set of protocols that enable third-party software applications to retrieve structured data from online platforms. Twitter streaming API provides streams of live feeds, and it is the most commonly used data source for Twitter research [12]. For this study, we ran Python scripts to access a 1% random sample through the sample function of Twitter streaming API [13]. From this sample, we retrieved MERS-related tweets using keywords, such as 'MERS' in Chinese, English, Indonesian/Malaysian (Bahasa Indonesia/Bahasa Malaysia), Japanese, Korean, and Thai. For details, see Online Supplementary Materials. The time frame was June 1 (the 13th day of the outbreak) to June 30, 2015 (4 days before the last case was confirmed), Seoul Time, i.e., Coordinated Universal Time (UTC) + 9. (The first of the 186 MERS cases was confirmed in Korea on May 20, 2015; by May 31, 23 cases were confirmed; the last four cases were confirmed in the first four days of July [10]. Our time frame captured the majority of the time span of the outbreak.)

We first performed a preliminary data analysis, using one week of data to fine-tune our data analysis plan, before we analyzed the full dataset.

After preliminary analyses, we focused our analysis on the corpora of 5 languages that returned the highest number of tweets: English ($N = 4024$), Indonesian ($N = 1256$), Japanese ($N = 1334$), Korean ($N = 21{,}823$) and Thai ($N = 2084$). We used Google Translate API to prepare preliminary English translations of the Indonesian, Japanese, Korean and Thai corpora. We then selected the posts with the highest frequency that contributed to 50% of the dataset and asked a native speaker or professional editor of each of the four languages to manually proofread the English translation.

Once we finalized the English translation, we performed the following analyses with the translated text.

General descriptive statistics

We calculated the basic descriptive statistics pertaining to the number of posts per user, retweets, hashtags, and embedded URL links. We estimated the Pearson's correlation between (i) the proportion of tweets with hashtags and the proportion of retweets, (ii) the proportion of retweets, and (iii) the proportion of tweets with embedded URL links.

User profile analysis

We analyzed a sample of user profiles and used open coding to annotate them to develop the following categories of users, as defined below:

- K-pop: Twitter profiles that were dedicated to Korean pop music (K-pop);
- Media: Twitter profiles of media organizations and journalists;
- Political: Twitter profiles of government agencies, political parties, politicians, and political activists;
- Medical: Twitter profiles of hospitals, clinics, health organizations, and individual healthcare workers;
- Others: Twitter profiles that do not belong to the four aforementioned categories.

The first coder randomly sampled a subset of Twitter user profiles (English, $n = 150$; other four languages, $n = 100$ each) and manually coded them by the aforementioned categories. The second coder randomly selected 50 user profiles (10 for each language) from the 550 manually coded profiles for double-coding that ensured a substantial inter-rater reliability (Cohen's kappa = 0.71). We performed Fisher's exact test to test for differences between groups.

We performed a χ^2 feature selection analysis to pick the keywords that are most unique to a specific corpus. χ^2 feature selection allows users to select subsets of features of data by analyzing the correlation between terms and categories, which is one of the most effective methods of feature selection [14]. We used the χ^2 feature selection algorithm (1) to evaluate how strong the evidence is for the hypothesis that the frequency of a word (or any other item) was different between the two corpora of tweets (such as Japanese versus non-Japanese), and (2) to rank the terms by χ^2 value that measured the extent of difference between the observed count and the expected count, assuming that the item's occurrence was independent of the corpus of tweets. A high χ^2 value indicates strong evidence for the existence of a difference between the item's observed frequency and its expected frequency, and therefore, its particularity to a specific corpus.

Retweets

Then we analyzed retweets to identify influential participation in the MERS discussion in each corpus. We first extracted all user names after 'RT@' from each data corpus, then we performed a frequency analysis to identify the most retweeted users. The number of retweets one receives is an important indicator of one's influence [15]. Thus the results from the retweet analysis were used to identify the most influential participants in the MERS discussion. We then selected the top 30 retweeted users from each corpus, and categorized them into the five aforementioned categories — K-pop, media, political, medical, Others — based on their Twitter user profile page. We performed Fisher's exact test to test their differences.

Statistical tests

Statistical analysis was performed with R 3.2.1 and RStudio 0.99.442, as well as with SAS 9.4. Significant level was defined *a priori* at the level of 0.05. For testing differences between categories of data presented in contingency tables, we use χ^2 test or Fisher's exact test where there were cells with $n < 5$. For larger than 2 by 2 tables, we either use the exact methods or compute p-values by Monte Carlo simulation as appropriate. For correlation between two continuous variables, we used Pearson's product—moment correlation and tested it against the null hypothesis of $r = 0$ with t-test.

Results

General descriptive statistics

Table 1 presents the descriptive statistics of our Twitter samples. We retrieved MERS-related Twitter data in five languages: Korean ($N = 21{,}823$), English ($N = 4024$), Thai ($N = 2084$), Japanese ($N = 1334$) and Indonesian ($N = 1256$) (Table 1). The mean number of posts per user

Table 1 Descriptive statistics of the five MERS-related Twitter corpora.

	Number of tweets (*N*)	Number of unique user	Median number of posts per user (IQR)	Mean number of posts per user	Retweets (% of *N*)	Posts containing hashtags (% of *N*)	Number of unique hashtags	Median number of posts per hashtag (IQR)	Mean number of posts per hashtag	Posts containing URLs (% of *N*)
Korean	21,823	14,646	1 (1, 1)	1.49	17,104 (78)	2791 (18)	1032	1 (1, 3.75)	2.7	9949 (46)
English	4024	3469	1 (1, 1)	1.16	1612 (40)	1212 (30)	593	1 (1, 2)	2.0	3136 (78)
Thai	2084	1991	1 (1, 1)	1.05	2008 (96)	1107 (53)	97	1 (1, 4.25)	11.4	347 (17)
Japanese	1334	1117	1 (1, 1)	1.19	693 (52)	424 (32)	291	1 (1, 2)	1.5	1086 (81)
Indonesian	1256	956	1 (1, 1)	1.31	98 (8)	215 (17)	143	1 (1, 2)	1.5	1142 (91)

IQR: interquartile range. We present the first quartile and the third quartile here.

ranged from 1.05 to 1.49. However, the distribution for the number of posts per user is highly skewed, with the median being 1 for all 5 corpora (with interquartile range: 1, 1). The percentage of retweets ranged from 8% (Indonesian) to 96% (Thai). Across the five language corpora, as the proportion of tweets with URLs increased, the proportion of retweets decreased ($r = -0.93$, $p = 0.02$).

User profile analysis

We found that there was statistically significant difference between the type of Twitter profiles of those who tweeted about MERS between the corpora of five different languages (Fisher's exact test, $p < 0.001$). We found that the percentage of users who were K-pop fans ranged from 4% in the Korean corpus to 70% in the Thai corpus (Table 2). For media, it ranged from 0% in the Thai corpus to 14% in the Indonesian corpus; political, 0% (Thai) to 19% (Japanese); medical, 0% (Thai) to 7% (English).

Keyword feature selection analysis

The χ^2 feature selection algorithm examined the particularity of discussion focus of each language corpus. Table 3 showed the top 20 most discriminative terms that appear in each corpus vis-à-vis the other corpora. The higher the rank (i.e., the greater the χ^2 value), the more unique this term was to the chosen language corpus. The exception was the word 'Korea' that appeared in the top 20 lists of more than one corpus.

Except for keywords that are apparently specific to this outbreak (i.e., 'MERS', 'Korea', 'Korean', and 'South Korea'), we found that there are unique keywords (when translated into English) used in MERS-related tweets that distinguish the corpora of five different languages. Both the Korean and English corpora featured keywords directly related to the outbreak. In the Japanese corpus, featured terms included 'Rakutenichiba' (a shopping website), 'rainy', 'mold', 'season', and 'acid' that are from commercial tweets used by online retailers to sell products (e.g., air purifiers and masks). Another feature of the Japanese corpus is the anti-Korean sentiment. The feature selection result of Thai tweets reveals that there were a few K-pop related posts, which can be informed by the keyword 'fan' and 'EXO'. Because there was a MERS case in Thailand (confirmed on June 17, 2015) [16], we also observed keywords 'withstand', 'tissue', and 'cough,' which all appeared in educational posts that educated the public about MERS virus. The most important feature of Indonesian corpus is the appearance of keywords related to Muslim communities. The keywords, including 'pilgrim', 'Umrah', and 'journey,' reflected the pilgrimage-related concerns of the Muslim majority in Indonesia.

Sources of retweets

For English, Japanese, Korean and Thai, the top 30 users' posts contributed to at least 20% of tweets in each of the data sets. For Indonesian, it is only 8% because there is only a very small portion of tweets were retweets in that corpus. Of the total 150 user profiles coded, 70 (46.7%) were media;

Table 2 Percentage of randomly sampled user profiles by language and by user categories.

		Categories of randomly sampled users (n, %)				
Language	Sample size	K-pop fan	Media	Political	Medical	Others
Korean	100	4 (4.0)	6 (6.0)	14 (14.0)	3 (3.0)	73 (73.0)
English	150	38 (25.3)	18 (12.0)	1 (0.7)	10 (6.7)	83 (55.3)
Thai	100	70 (70.0)	0 (0.0)	0 (0.0)	0 (0.0)	30 (30.0)
Japanese	100	18 (18.0)	3 (3.0)	19 (19.0)	1 (1.0)	59 (59.0)
Indonesian	100	7 (7.0)	14 (14.0)	4 (4.0)	2 (2.0)	73 (73.0)

29 (19.3%) were K-pop fans; 7 (4.7%) were political; 9 (6%) were medical; and 35 (23.3%) were categorized as 'Others' (Table 4). We found that the types of the top 30 Twitter profiles that received the most retweets for their MERS-related tweets differed between the corpora of five different languages (Fisher's exact test, $p < 0.001$).

Discussion

We analyzed five corpora (English, Indonesian, Japanese, Korean and Thai) of MERS-related Twitter data from June 1 to June 30, 2015. We identified differences between the five corpora in terms of general user profiles, keywords, and types of the top 30 Twitter profiles that received most retweets.

Our findings suggest that the reactions to the 2015 MERS outbreak in South Korea differed between Twitter users of different languages. This observation further suggests that results of Twitter studies based on an English-only corpus may not always be generalizable to non-English-speaking Twitter users. Recent research identified significant correlation between the normalized volume of tweets with one of three Korean MERS-related keywords with number of laboratory-confirmed MERS cases in the outbreak [17]. Our study extended the existing literature on Twitter and MERS in Korea by going beyond the English and Korean languages by including tweets in Indonesian, Japanese and Thai.

Our study has several implications for public health practitioners. First, different linguo-cultural communities on Twitter may react to an outbreak differently. An understanding of the social media culture of a specific linguistic community will contribute to the success of any effective social media public health communications directed at the said community.

Second, machine translation (such as Google Translate) with a reasonable level of human involvement can enable us to monitor Twitter reaction in foreign languages on social media. This is useful for both agencies with global health responsibilities (such as the World Health Organization),

Table 3 Top 20 keywords for each language corpus, identified by χ^2 feature selection method.

	Korean	Japanese	Thai	Indonesian	English
1	Korea	Japan	Thai	Indonesia	Korea
2	Outbreak	Rakutenichiba[b]	Postpone	Korea	Outbreak
3	Case	Mutual	Marriage	Beware	Case
4	Patient	Maintain	Scarier	Pilgrim	Reuter
5	Hospital	Rainy	Easier	Attack	News
6	Park Wonsoon[a]	Korea	Withstand	Aware	Report
7	Government	Mold	Spring-news	Victim	Death
8	News	Bulletin	Develop	Expel	South Korea
9	Korean	Yahoo	Pound	Alert	Rise
10	Thailand	Infect	Strain	News	Patient
11	Travel	Person	Picture	Anticipate	Government
12	Reuters	Affair	Anthem	Complex	Contain
13	President	Travel	Embassy	Consulate	BTS[e]
14	Death	NHK[c]	Heavily	Hospital	Infection
15	Thai	Season	Language	K-pop	People
16	Virus	Labor	Tissue	Umrah	Park
17	Report	Acid	Kidney	Toddler	Virus
18	Middle East	Perish	EXO[d]	Case	Aid
19	Postpone	Anti-Korean	Fan	Health	Hospital
20	Center	News	Cough	Journey	Fifth

[a] Park Wonsoon: The mayor of Seoul as of June 2015.
[b] Rakutenichiba: a Japanese online shopping site.
[c] NHK: Nippon Hoso Kyokai (Japan Broadcasting Corporation), the national public broadcasting organization.
[d] EXO: a Korean pop band.
[e] BTS: a Korean pop band.

Table 4 Top 30 sources of retweets by language and by categories.

		Categories of top 30 sources of retweets (*n*, %)				
Language	Total sample	K-pop fan	Media	Political	Medical	Others
Korean	30	1 (3.3)	13 (43.3)	4 (13.3)	0 (0.0)	12 (40.0)
English	30	5 (16.7)	20 (66.7)	0 (0.0)	1 (3.3)	4 (13.3)
Thai	30	14 (46.7)	8 (26.7)	0 (0.0)	2 (6.7)	6 (20.0)
Japanese	30	4 (13.3)	10 (33.3)	5 (16.7)	3 (10.0)	8 (26.7)
Indonesian	30	5 (16.7)	19 (63.3)	0 (0.0)	1 (3.3)	5 (16.7)
	150	29 (19.3)	70 (46.7)	9 (6.0)	7 (4.6)	35 (23.3)

and agencies that serve populations that speak multiple languages (such as the Centers for Disease Control and Prevention in the US, where there is a significant Spanish-speaking minority).

Third, the observation that Korean pop culture plays an important role among Thai Twitter users' communication about MERS echoes Brown et al.'s argument that the responsible and creative use of pop culture may help public health professionals reach hard-to-reach communities via social media with health information in an outbreak emergency response [18]. The K-pop cultural connection helped facilitate the circulation of MERS-related information in Thailand, where there was also one confirmed MERS case imported from the Middle East [16].

There are some limitations. First, Twitter users may not be representative of the average speaker of a given language. For example, while Twitter enjoys popularity among English speakers, it may be a minority interest among Chinese speakers [8]. Second, linguo-cultural communities may span across different countries. The global diaspora of Indonesian, Japanese, Korean and Thai speakers span across continents. For example, there is a sizeable minority of Korean speakers in Los Angeles and Atlanta, USA. Similarly, English is now a global language and is spoken as a second language in many countries. Non-native speakers may tweet in English. Third, the Indonesian and Malaysian languages are very similar and mutually intelligible. Our corpus of Indonesian may include tweets in Malay. We used the original Twitter language designation because we offered no better solution than Twitter's own language detection algorithm. Fourth, our 1% random sample of tweets, while representative, was a small sample.

To conclude, we observed significant differences in MERS-related Twitter activity between the corpora of Korean, English, Thai, Japanese, and Indonesian languages in terms of users, and keywords. Our results suggest that findings based on analyzing English tweets alone may not be generalizable to Twitter users that tweet in other languages. Analysis of outbreak-related Twitter corpus specific to a given language will help national public health agencies across the globe understand their own communities' Twitter cultures to develop culture-specific Twitter health communication strategy.

Ethics

Our study is approved by Georgia Southern University's Institution Review Board (H15083) under the B2 exempt category.

Authorship statement

ICHF, JZ, KWF and ZTHT conceived and designed the project. ZL and ZTHT collected the Twitter data. ICHF and JZ performed preliminary data analysis on a subset of the data. JZ performed the final data analysis on the full dataset under the guidance of ICHF and KWF. ICHF performed data fitting (power-law distribution) and some statistical tests. ICHF created Supplemental Figures 1 and 3. JY performed the ANCOVA analysis and created Supplemental Figures 2 and 4. CHC, HL and KWF provided consultation on R programming to and shared their R codes with ICHF and JZ. JY and KWF provided statistical consultation to ICHF regarding data fitting and statistical analysis. ICHF and JZ co-wrote the early drafts of this manuscript. All co-authors edited the manuscript with intellectual inputs and approved its final version for submission to the journal.

Conflict of interest

We declare that we have no competing interests.

Funding

We do not receive any external funding for this project.

Provenance and peer review

Not commissioned; externally peer reviewed.

Disclaimer

The opinions expressed in this paper do not represent the official positions of the Centers for Disease Control and Prevention or the United States Government.

Acknowledgment

The authors thank Fiona Suwana, Yuko Clay, Juwon Park, Yizhou (May) Shi for manual editing of the Google English translation of Indonesian, Japanese, Korean and Thai tweets, respectively. ICHF and JZ thank the Journalism and Media Studies Centre, the University of Hong Kong, for accommodating their academic visits during summer 2015 when a significant part of this research project was done. ICHF and ZTHT received salary support from the Centers for

Disease Control and Prevention (CDC) (#15IPA1509134 and #16IPA1619505). This paper is not part of their CDC-supported research. The CDC has no role in its study design, data analysis, writing or its submission for publication.

Appendix A. Supplementary data

Supplementary data related to this article can be found at http://dx.doi.org/10.1016/j.idh.2017.08.005.

References

[1] Fung IC-H, Fu K-W, Ying Y, Schaible B, Hao Yi, Chan C-H, et al. Chinese social media reaction to the MERS-CoV and avian influenza A(H7N9) outbreaks. Infect Dis Poverty 2013;2:31.

[2] Lupton D. Medicine as culture: illness, disease and the body in western societies. London: Sage Publications; 2003.

[3] Statistica. Number of monthly active Twitter users worldwide from 1st quarter 2010 to 2nd quarter 2017 (in millions). Available from: https://www.statista.com/statistics/282087/number-of-monthly-active-twitter-users/; 2017.

[4] Statistica. Number of active Twitter users in leading markets as of May 2016 (in millions). Available from: https://www.statista.com/statistics/242606/number-of-active-twitter-users-in-selected-countries/; 2016.

[5] Twitter. Company. Available from: https://about.twitter.com/company; 2016.

[6] Fung IC-H, Duke CH, Finch KC, Snook KR, Tseng PL, Hernandez AC, et al. Ebola virus disease and social media: a systematic review. Am J Infect Control 2016;44:1660—71.

[7] Fu KW, Liang H, Saroha N, Tse ZT, Ip P, Fung IC. How people react to Zika virus outbreaks on Twitter? A computational content analysis. Am J Infect Control 2016;44:1700—2.

[8] Mocanu D, Baronchelli A, Perra N, Goncalves B, Zhang Q, Vespignani A. The Twitter of Babel: mapping world languages through microblogging platforms. PLoS One 2013;8:e61981.

[9] Seidlhofer B. Anglophone-centric attitudes and the globalization of English. J Eng Lingua Franca 2012;1:393—407.

[10] Korea Centers for Disease Control and Prevention. Middle East respiratory syndrome coronavirus outbreak in the Republic of Korea, 2015. Osong Public Health Res Perspect 2015;6:269—78.

[11] Twitter. Twitter developer documentation: streaming APIs. Available from: https://dev.twitter.com/streaming/overview; 2017.

[12] Gaffney D, Puschmann C. Data collection on Twitter. In: Weller K, Bruns A, Burgess J, Mahrt M, Puschmann C, editors. Twitter and society. New York: Peter Lang; 2014.

[13] Twitter. Twitter developer documentation: GET statuses/sample. Available from: https://dev.twitter.com/streaming/reference/get/statuses/sample; 2017.

[14] Yang Y, Pedersen J. A comparative study on feature selection in text categorization. In: Proceedings of the international conference on machine learning; 1997. p. 412—20.

[15] Bruns A, Stieglitz S. Towards more systematic Twitter analysis: metrics for tweeting activities. Int J Soc Res Methodol 2013;16:91—108.

[16] Wiboonchutikul S, Manosuthi W, Likanonsakul S, Sangsajja C, Kongsanan P, Nitiyanontakij R, et al. Lack of transmission among healthcare workers in contact with a case of Middle East respiratory syndrome coronavirus infection in Thailand. Antimicrob Resist Infect Control 2016;5:21.

[17] Shin SY, Seo DW, An J, Kwak H, Kim SH, Gwack J, et al. High correlation of Middle East respiratory syndrome spread with Google search and Twitter trends in Korea. Sci Rep 2016;6:32920.

[18] Brown B, Nasiruddin M, Dao A, Halabi M. Responsible use of pop culture and communication in the face of Ebola virus. PLoS Negl Trop Dis 2015;9:e0003890.