

Social Media's Initial Reaction to Information and Misinformation on Ebola, August 2014: Facts and Rumors

ISAAC CHUN-HAI FUNG, PhD^a
KING-WA FU, PhD^b
CHUNG-HONG CHAN, MSc^b
BENEDICT SHING BUN CHAN,
PhD^c
CHI-NGAI CHEUNG, MPhil^d
THOMAS ABRAHAM, MA^b
ZION TSZ HO TSE, PhD^e

ABSTRACT

Objective. We analyzed misinformation about Ebola circulating on Twitter and Sina Weibo, the leading Chinese microblog platform, at the outset of the global response to the 2014–2015 Ebola epidemic to help public health agencies develop their social media communication strategies.

Methods. We retrieved Twitter and Sina Weibo data created within 24 hours of the World Health Organization announcement of a Public Health Emergency of International Concern (Batch 1 from August 8, 2014, 06:50:00 Greenwich Mean Time [GMT] to August 9, 2014, 06:49:59 GMT) and seven days later (Batch 2 from August 15, 2014, 06:50:00 GMT to August 16, 2014, 06:49:59 GMT). We obtained and analyzed a 1% random sample of tweets containing the keyword Ebola. We retrieved all Sina Weibo posts with Chinese keywords for Ebola for analysis. We analyzed changes in frequencies of keywords, hashtags, and Web links using relative risk (RR) and χ^2 feature selection algorithm. We identified misinformation by manual coding and categorizing randomly selected sub-datasets.

Results. We identified two speculative treatments (i.e., bathing in or drinking saltwater and ingestion of Nano Silver, an experimental drug) in our analysis of changes in frequencies of keywords and hashtags. Saltwater was speculated to be protective against Ebola in Batch 1 tweets but their mentions decreased in Batch 2 (RR=0.11 for "salt" and RR=0.14 for "water"). Nano Silver mentions were higher in Batch 2 than in Batch 1 (RR=10.5). In our manually coded samples, Ebola-related misinformation constituted about 2% of Twitter and Sina Weibo content. A range of 36%–58% of the posts were news about the Ebola outbreak and 19%–24% of the posts were health information and responses to misinformation in both batches. In Batch 2, 43% of Chinese microblogs focused on the Chinese government sending medical assistance to Guinea.

Conclusion. Misinformation about Ebola was circulated at a very low level globally in social media in either batch. Qualitative and quantitative analyses of social media posts can provide relevant information to public health agencies during emergency responses.

^aGeorgia Southern University, Jiann-Ping Hsu College of Public Health, Department of Epidemiology, Statesboro, GA

^bThe University of Hong Kong, Journalism and Media Studies Centre, Hong Kong

^cHong Kong Baptist University, Department of Religion and Philosophy, Hong Kong

^dEmory University, Department of Psychology, Atlanta, GA

^eUniversity of Georgia, College of Engineering, Athens, GA

Address correspondence to: Isaac Chun-Hai Fung, PhD, Georgia Southern University, Jiann-Ping Hsu College of Public Health, Department of Epidemiology, PO Box 8015, Statesboro, GA 30460; tel. 912-478-5079; fax 912-478-0171; e-mail <cfung@georgiasouthern.edu>.

©2016 Association of Schools and Programs of Public Health

Communicating scientifically accurate information about an outbreak is important, because an informed public will likely be less susceptible to misinformation that could hinder outbreak control.¹ Although social media have been used by public health agencies to communicate disease prevention information,^{2–4} rumors and alternate understandings of disease can circulate. Anecdotal evidence suggests that Twitter might have played a role in Nigeria's efforts to control Ebola⁵ at the outset of the 2014 Ebola outbreak, but the World Health Organization (WHO) noted rumors circulating on social media claiming that certain products or practices could prevent or cure Ebola virus disease.⁶ A 2014 study found that 55% of English-language tweets from Guinea, Liberia, and Nigeria during September 1–7, 2014, using the terms “Ebola” and “prevention” or “cure” contained medical misinformation.⁷

We analyzed information and misinformation about Ebola on Twitter—the world's largest microblogging service—and Sina Weibo—the leading microblogging platform in China—shortly after the WHO declaration of the Ebola outbreak as a Public Health Emergency of International Concern (PHEIC) on August 8, 2014.⁸ Our study focused on microblogging, an Internet-based self-publishing application that enables people to share user-created content or republish others' messages online. It has played a critical role in mass communication during crises, such as natural disasters.^{9–11}

Microblog users' response was part of the public's reaction to the global threat of Ebola. Most people had insufficient knowledge about the disease and limited time and resources to access additional information sources to understand it.^{12,13} We used Twitter as a representative social media platform. We included Sina Weibo because Twitter is blocked in China, requiring microblog users in China to use alternative platforms.

Our primary research questions were (1) What Ebola-related information and misinformation (and their proportions) was circulated on two popular microblogging platforms (Twitter and Sina Weibo) in the two most commonly spoken languages (English and Chinese) on the day of the PHEIC announcement? and (2) What changes could be observed a week later? Our analysis may aid public health agencies in developing their social media communication strategies.

METHODS

Study design

We collected microblog data from Twitter and Sina Weibo in two batches to document changes in Ebola-related microblog content one week after the WHO PHEIC announcement. Batch 1 was collected during

the first 24 hours after the WHO PHEIC announcement¹² (from August 8, 2014, 06:50:00 Greenwich Mean Time [GMT] to August 9, 2014, 06:49:59 GMT); Batch 2 was collected during a 24-hour period seven days later (from August 15, 2014, 06:50:00 GMT to August 16, 2014, 06:49:59 GMT).

Twitter data retrieval

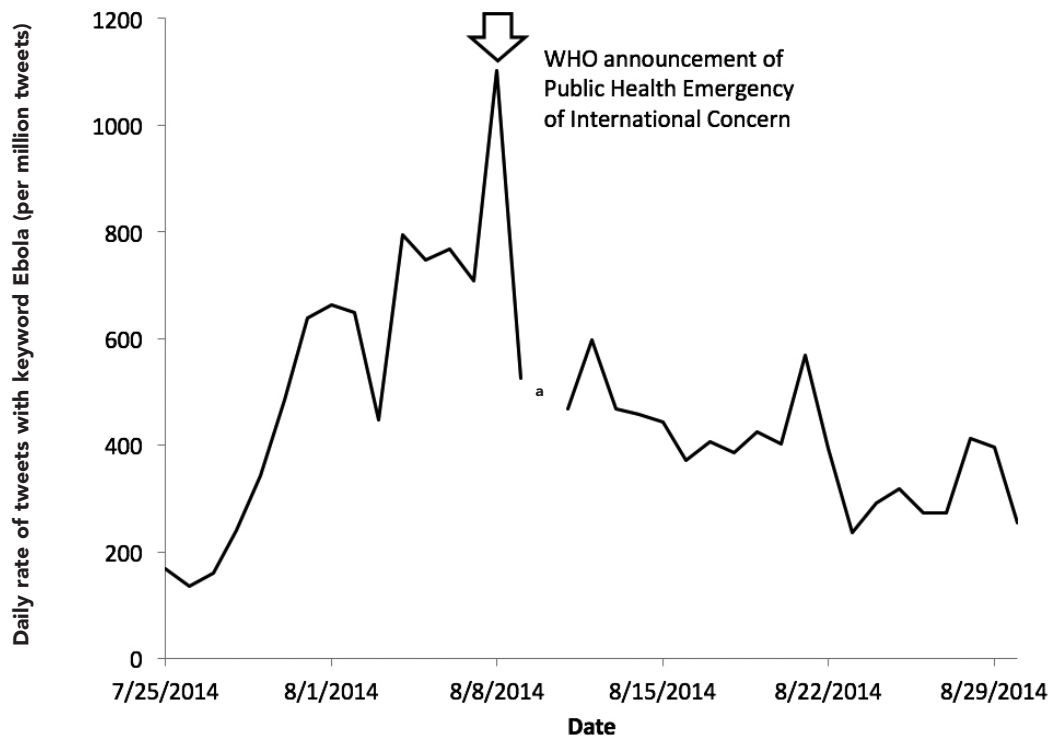
Twitter had 271 million monthly active users worldwide in 2014.¹³ Tweets are seen as digital footprints for monitoring the public's health-related responses and behaviors.^{14,15} We retrieved Twitter users' publicly available tweets. We applied Twitter streaming Application Programming Interface (API) to sample Twitter data for monitoring the Ebola outbreak (details available upon request). We retrieved a random sample of 1% of all publicly posted tweets: (sample size: $n=4,366,946$ tweets from Batch 1 and $n=4,305,841$ tweets from Batch 2). Of these, 4,844 tweets during Batch 1 and 2,001 tweets during Batch 2 (1,109 and 465 per one million tweets, respectively) contained the keyword “Ebola” (Figure).

Chinese microblog data retrieval

Sina Weibo had 156.5 million monthly active users as of June 2014¹⁶ and is often used for digital epidemiology studies on Chinese social media.¹⁷ Despite the state's control of political information, Chinese online users can use Sina Weibo to speak with some autonomy on public health affairs.

Because Sina Weibo currently does not provide a streaming API service similar to that of Twitter, Chinese Ebola-related microblogs were obtained via Sina Weibo's Internet search engine. Two Chinese terms for Ebola, 伊波拉 and 埃博拉, were entered into the search engine. The results were then computationally captured page by page by a script (developed by one of the authors) based on R version 3.0.2.¹⁸ The script was programmed to run every 10 minutes during the Batch 1 and Batch 2 periods and was executed by two computers located separately in Hong Kong and Athens, Georgia, for data redundancy management. When data collection was completed, items with duplicated identity codes were discarded. Of the 7,645 microblog posts in Batch 1, 219 posts created outside the specified time frame and 49 posts containing the Chinese term for “Ebola” in their username but not in the body of the posts were excluded, leaving 7,377 posts for further text processing. Of the 3,416 microblog posts in Batch 2, 64 posts containing the Chinese term for “Ebola” in their username but not in the body of the posts were excluded, leaving 3,352 posts for further text processing. Finally, we parsed the text of the col-

Figure. Incidence rate of Twitter tweets with keyword “Ebola” (count per million per day) in a 1% random sample of the Twitter universe, July 25–August 30, 2014



^aTwitter data retrieval on August 10, 2014, was incomplete because of server problems; as such, Twitter data were not included and the line does not show these data.

lected microblog messages and the time of posting and recorded them in a comma-separated value (CSV) file.

Text processing

The retrieved Twitter tweets were classified into English and non-English languages by using R;¹⁹ only English-language tweets were analyzed. Using R's text mining package,²⁰ tweets were then stemmed by stemmer (e.g., reducing words to their roots, such as “tables” to “table”) and tokenized for keyword analysis. Tokenization is the process of segmenting a sentence into different units of meaning (words). Because of the linguistic difference between English-language tweets and Chinese microblogs, the Chinese microblogs were processed differently. Because there is no space in the Chinese syntax to separate words in a sentence as in English, Chinese text in microblogs was segmented into phrases using Viterbi algorithm²¹ implemented in the *Jieba* toolkit.^{22,23} For example, an original Chinese sentence such as

埃博拉病毒病（以往称作埃博拉出血热）是由丝状病毒科的埃博拉病毒导致的一种严重且往往致命的疾病，死亡率高达90%。

[English translation: Ebola virus disease (formerly known as Ebola hemorrhagic fever) is a severe, often fatal illness, with a death rate of up to 90% caused by Ebola virus, a member of the filovirus family.]

was segmented into phrases by the Viterbi algorithm as

埃博拉/病毒/病/（/以往/称作/埃博拉/出血热/）/是/由/丝状/病毒科/的/埃博拉/病毒/导致/的/一种/严重/且/往往/致命/的/疾病/，/死亡率/高达/90/%。

A word-by-word translation in English would be: Ebola / virus / disease / formerly / called / Ebola / hemorrhagic fever / is / from / filial / virus /'s / Ebola virus / cause /'s/ one kind / severe / and / eventually / fatal /'s / disease / death rate/ as high as / 90 / %.

Candidate keywords for subsequent analysis were all unique keywords in the tokenized English-language tweets and segmented Chinese microblogs with more

than three occurrences in the entire collection of tweets and microblogs. We extracted hashtags and Web links from the body of the tweets and the Chinese microblogs. The extracted Web links were mostly shortened URLs and were resolved by the *cURL* program²⁴ to obtain the full Web links and domain names. We removed stop words, punctuation, emoticons, and special symbols used in microblogging (e.g., RT and @author_name) in the English-language and Chinese microblogs. We recorded the contents and time of posting for each microblog post in a CSV file.

Analysis of changing relative frequencies

We analyzed seven-day changes in the relative frequencies of keywords, hashtags, and domain names of the shared Web links. The unit of analysis was a keyword, a hashtag, or a domain name. “Trending” signifies increasing usage of the item, and “fading” signifies declining usage. We evaluated the relative frequency of each item in Batch 1 and Batch 2 using the relative risk (RR) and χ^2 feature selection algorithm.²⁵

The χ^2 feature selection algorithm (1) evaluated the strength of evidence for the hypothesis that the frequency of an item between the two batches of microblogs was different and (2) ranked terms by the χ^2 value that measured how much the observed count was different from the expected count, assuming that occurrence of the item was independent of the batch of microblogs. High χ^2 values indicate strong evidence for the existence of a difference between the observed value and expected value.²⁶ Such measurements did not account for the direction of the difference (i.e., an item with a high χ^2 value may have a higher frequency in Batch 2 than in Batch 1, or vice versa). We used RR to supplement the χ^2 feature selection algorithm to indicate the direction of the relative frequency. The RR for an item i was calculated as:

$$RR_i = (P_{i, \text{Batch 2}}) / (P_{i, \text{Batch 1}})$$

$P_{i, \text{Batch 1}}$ denoted the probability of tweets with item i in Batch 1; $P_{i, \text{Batch 2}}$ denoted the probability of tweets with item i in Batch 2. A 0.5 was added to both its denominator and numerator to correct for zero frequencies. In this analysis, trending items were those with a high χ^2 value and an $RR > 1$, while fading items were those with a high χ^2 value and an $RR \leq 1$.

Manual coding and categorization of microblog contents

We manually categorized microblog content under different themes to identify the information and misinformation contained therein. For manual coding, we randomly selected 5%–7% of the de-identified social

media posts. In Microsoft® Excel® spreadsheets, for each Twitter tweet or Sina Weibo post, we assigned a random number between 0 and 100 inclusive (=RANDBETWEEN[0,100]). If the random number was ≤ 5 , the microblog post was selected for manual coding. Because the proportion of random numbers generated that were ≤ 5 was not the same for each dataset, the manually coded dataset was 5%–7% of the original datasets: Twitter: Batch 1: $n=299/4,844$ (6.2%); Batch 2: $n=116/2,001$ (5.8%); Sina Weibo: Batch 1: $n=469/7,645$ (6.1%); Batch 2: $n=207/3,416$ (6.1%). After random selection, 17 of the 469 randomly selected Sina Weibo posts in Batch 1 were excluded because they were outside the time frame, leaving 452 manually coded posts.

Our social media samples were first coded (i.e., categorized under different themes) by at least one coder and then recoded by the first author, who made the final coding decision. To assess reliability, a second coder coded a randomly selected 10% sample of our manually coded samples. Comparing the code between the second coder and the first author, the interrater agreement was moderate for Twitter data (Cohen’s $\kappa=0.58$ for Batch 1 and $\kappa=0.56$ for Batch 2) and substantial for Sina Weibo data (Cohen’s $\kappa=0.66$ for Batch 1 and $\kappa=0.78$ for Batch 2).²⁷

Within each selected sample, the microblog posts were first categorized into English/Chinese posts and posts that were not English or Chinese. For Twitter data, the non-English, non-Chinese tweets (54 in Batch 1 and 29 in Batch 2) were excluded from further analysis, leaving 245 tweets in Batch 1 and 87 posts in Batch 2 for analysis. We found no significant difference in the proportion of English-language tweets between Batch 1 and Batch 2 ($p=0.15$). We manually categorized English/Chinese posts in our selected samples according to the following scheme (derived from a previously published scheme)²⁸: news of the Ebola outbreak or cases; news of travel bans, border blockades, flight route suspensions, sports game bans, and travel advice; health education and information; alternative health information; responses to alternative health information; advertisement and entertainment; social issues; and others.

In our analysis of Sina Weibo samples, we created two categories for posts related to the Chinese medical team’s departure for Guinea and for posts that reported news of a Sina Weibo user who allegedly spread rumors about a “suspected case in a hospital in Shanghai” and was detained by Chinese police. These news items were unique to Sina Weibo content. We performed Fisher’s exact test and χ^2 tests as appropriate to compare the number of posts in each category between the two

batches for Twitter and Sina Weibo, respectively. The Twitter and Sina Weibo data were anonymized and de-identified prior to analysis.

RESULTS

Descriptive analysis

Key terms used in the WHO's PHEIC announcement were among the top 20 keywords list of the English and Chinese microblogs in Batch 1: "Ebola," "virus," "outbreak/epidemic," "public health," and "emergency" (Table 1).

Analysis of changing relative frequencies

Terms related to the WHO's PHEIC announcement, including "emergency event" ($\chi^2=458.9$, $RR=0.01$), "public health" ($\chi^2=371.0$, $RR=0.10$), or "announced" ($\chi^2=232.0$, $RR=0.12$), were fading in Batch 2 on Sina Weibo, while the term "declare" was fading in Batch 2 on Twitter (Table 2).

Two rumors featured in numerous posts, but both were confined to Twitter and did not appear on Sina Weibo. Saltwater appeared in Batch 1 tweets but faded in Batch 2 (salt: $\chi^2=63.9$, $RR=0.11$; water: $\chi^2=39.7$, $RR=0.14$). An experimental drug known as Nano Silver, which was backed by the Nigerian Ministry of Health but had no scientific evidence on efficacy, was on the top 10 trending Twitter hashtag list ($\chi^2=1.7$, $RR=10.5$) (Table 2).

Tweets reading "Ebola may be vastly underestimated," reflecting a WHO assessment, made "underestimate" the top trending term on Twitter in Batch 2. In Chinese microblogs, the term "assistance to Africa" ranked at the top of the trending list in Batch 2 (Table 2).

Manual coding and categorization of microblog contents

Twitter. We found a significant difference (Fisher's exact test, $p=0.01$) in content categories between Batch 1 and Batch 2. Of 245 tweets in Batch 1 and 87 tweets in Batch 2, alternative health information (i.e., information that is not in line with current scientific understanding of Ebola and its prevention and control) accounted for six tweets in Batch 1 and two tweets in Batch 2. We found a similar percentage of tweets on health education and information (i.e., information that is in line with current scientific understanding of Ebola and its prevention and control) (Batch 1: $n=26$, 10.6%; Batch 2: $n=10$, 11.5%) and tweets in response to alternative health information (Batch 1: $n=32$, 13.1%; Batch 2: $n=10$, 11.5%) (Table 3).

Sina Weibo. We observed a significant difference in the manually coded categories between randomly selected Sina Weibo posts from Batch 1 and Batch 2 (Fisher's exact test, $p<0.001$). Of 452 posts analyzed in Batch 1 and 207 posts analyzed in Batch 2, alternative health information accounted for 11 posts in Batch 1 and three posts in Batch 2. Percentage of posts about health education and information were similar between the two batches (Batch 1: $n=55$, 12.2%; Batch 2: $n=27$, 13.0%). Responses to alternative health information accounted for 43 (9.5%) posts in Batch 1 and 13 (6.3%) posts in Batch 2 (Table 3).

Alternative health information as observed through manual coding

Twitter. In the manually coded subset, alternative health information about Ebola in Batch 1 included two posts on "bathing in salt water and then drinking it"; one post that the Ebola virus came from space; one post that "crystal meth can cure Ebola"; one post that Ebola arose from human cannibalism; and one post to "tie a palm leaf and a red cloth round your head and your waist, and then dance around any banana tree" to stop Ebola. Alternative health information in Batch 2 included two posts on news about the use of Nano Silver as an Ebola treatment in Nigeria.

Sina Weibo. In the manually coded subset, alternative health information about Ebola on Sina Weibo in Batch 1 included three posts with scientific comments that contained mistakes (e.g., "from its discovery to today, Ebola had only caused havoc in West Africa"), two posts that advocated homeopathy, and six posts that said traditional Chinese medicine could be used to treat or prevent Ebola. Two posts advocating traditional Chinese medicine criticized the Western medical idea of Ebola virus causing the Ebola virus disease. That particular comment drew 25 posts containing criticism and sarcastic comments from Sina Weibo users. In Batch 2, examples included two posts on traditional Chinese medicine and one post that smoking tobacco can prevent Ebola infection.

DISCUSSION

Our finding that about 2% of microblogs contained alternative health information in August 2014 contrasts with a September 2014 study that focused on English-language tweets from Guinea, Liberia, and Nigeria and found that 55% of Ebola-related tweets contained medical misinformation.⁷ Whereas that study focused on three African countries affected by the outbreak, we sampled our microblog contents without geographical

Table 1. Top 20 most frequent keywords, top 10 most frequent Web link URL domains, and top 10 most frequent hashtags identified in two cross-sectional samples of Twitter tweets and Chinese microblogs (Sina Weibo) on Ebola, August 8–9, 2014, and August 15–16, 2014^a

Variable	Chinese microblogs (Sina Weibo)			
	Twitter		Batch 2: August 15–16, 2014	
	Batch 1: August 8–9, 2014	Batch 2: August 15–16, 2014	Batch 1: August 8–9, 2014	Batch 2: August 15–16, 2014
Number of Ebola posts retrieved (raw data)	4,844	2,001	7,645	3,416
Number of relevant posts analyzed ^b	2,132	820	7,377	3,352
Number of posts with hashtag (percent of analyzed posts)	601 (28.2)	306 (37.3)	629 (8.5)	1,394 (41.6)
Number of posts with URL (percent of analyzed posts)	1,127 (52.9)	523 (63.8)	1,459 (19.8)	470 (14.0)
Top 20 most frequent keywords ^c	ebola emerge health declare intern outbreak virus salt africa nigeria water spread world people amp ^e west patient prevent via public facebook.com nytimes.com akpraise.com bbc.co.uk edition.cnn.com reuters.com lindaikeji.blogspot.hk foxnews.com news.yahoo.com time.com	ebola outbreak say africa virus drug west people can health will doctor nigeria month underestimate experiment case vast via crisis nytimes.com usaheadlines.today edition.cnn.com bbc.co.uk foreignpolicy.com news.yahoo.com thejournal.ie todayng usatoday.com lindaikeji.blogspot.hk	伊波拉 (ebola) ^d 病毒 (virus) 疫情 (epidemic) 国际 (international) 链接 (links) 网页 (web page) 中国 (china) 西非 (west africa) 公共卫生 (public health) 分享 (share it) 紧急事件 (emergency event) 非洲 (africa) 宣布 (announced) 世界卫生组织 (WHO) 美国 (USA) 国家 (state) 肆虐 (raging) 抗体 (antibodies) 死亡 (death) 成为 (become) ifeng.com weibo.com sina.com.cn qq.com xinhaunet.com 163.com sohu.com infzm.com chinanews.com huanqiu.com	中国 (china) 伊波拉 (ebola) 非洲 (africa) 这片 (this piece) 病毒 (virus) 国家 (state) 医生 (doctors) 致敬 (tribute) 医疗队 (medical teams) 援非 (assistance to africa) 阻击 (check) 年前 (ago) 飞机 (aircraft) 派出 (sent) 几个 (several) 白衣天使 (angels in white) 下来 (down) 家人 (family members) 遥远 (distant) 阿尔及利亚 (algeria) qq.com weibo.com xinhaunet.com sina.com.cn ifeng.com mydrivers.com sina.cn 163.com gov.cn guokr.com
Top 10 most frequent Web link URL domains				

continued on p. 467

Table 1 (continued). Top 20 most frequent keywords, top 10 most frequent Web link URL domains, and top 10 most frequent hashtags identified in two cross-sectional samples of Twitter tweets and Chinese microblogs (Sina Weibo) on Ebola, August 8–9, 2014, and August 15–16, 2014^a

Variable	Chinese microblogs (Sina Weibo)			
	Twitter		Batch 1: August 8–9, 2014	Batch 2: August 15–16, 2014
Top 10 most frequent hashtags	ebola africa cdcchat news nigeria ebolaoutbreak factsonEbola breaking morelikely ebolafact	ebola usaheadlines nigeria breaking africa news ebolaoutbreak health ebolakenya liberia	Batch 1: August 8–9, 2014 伊波拉病毒 (Ebola virus) 新浪新闻有奖分享 (Sina news to share with prizes) 我有话说 (I have something to say) 维和前线 (peacekeeping forefront) 我在看新闻 (I was watching the news) 伊波拉 (Ebola) 索命病毒~伊波拉 (deadly virus~Ebola) 南方周末•新闻 (NFZM • news) 粉碎谣言 (dispelled the rumors) 第1新闻 (first news)	Batch 2: August 15–16, 2014 阻击伊波拉 (blocking ebola) 伊波拉病毒 (Ebola virus) 我在看新闻 (I was watching the news) 新浪新闻有奖分享 (Sina news to share with prizes) 伊波拉凶猛 (Ebola ferocious) 资讯提醒 (reminder) 伊波拉 (Ebola) 我在看今日话题 (reading today's topic) 我有话说 (I have something to say) 伊波拉病毒的预防 (Ebola virus prevention)

^aWe expanded the shortened URL Web links contained in the text of the tweets and Chinese microblogs into their original URL Web links and then identified their domains. Hashtags are used in Twitter and Chinese microblogs for a variety of reasons, such as emphasizing the theme of a tweet. For example, #cdcchat is a hashtag used by the Centers for Disease Control and Prevention (CDC) and its followers for their CDC chat activity on Twitter.

^bOf 4,844 and 2,001 tweets obtained in Batch 1 and Batch 2, respectively, only 2,132 and 820 tweets, respectively, were classified as English language by the language detection algorithm. Of 7,645 Chinese microblog posts obtained in Batch 1, 219 were excluded because they were posted outside the specified time frame, and 49 were excluded because they contained the Chinese term for "Ebola" in the username but not in the body of the Sina Weibo post, leaving 7,377 posts for analysis. Of 3,416 Chinese microblog posts in Batch 2, 64 were excluded because they contained the Chinese term for "Ebola" in the username but not in the body of the Sina Weibo post, leaving 3,352 posts for analysis.

^cTerms with the same stem (i.e., root word) were grouped under one category showing their infinite form. For example: the words "spread," "spreads," and "spreading" shared the same stem and were grouped under "spread."

^dTo unify the two Chinese translations of the keyword "Ebola," we converted all 埃博拉 into 伊波拉 during our analysis.

^eRefers to either a "&" (ampersand) sign or an abbreviation such as "amp up" or "accelerated manufacture of pharmaceuticals."

Table 2. Trending and fading analysis of keywords, Web link URL domains, and hashtags identified in two cross-sectional samples of Twitter tweets and Chinese microblogs (Sina Weibo) on Ebola, August 8–9, 2014, and August 15–16, 2014^a

Keywords	Chinese microblogs (Sina Weibo)											
	Twitter ^b				Trending				Fading			
	χ^2	RR	Keywords	RR	χ^2	RR	Keywords	RR	χ^2	RR	Keywords	RR
underestimate	108.1	77.10	emerge	0.04	174.6	0.04	assistance to Africa	186.38	2,754.8	458.9	emergency event	0.01
vast	101.8	210.52	declare	0.03	118.8	0.03	sent	945.56	2,613.3	435.9	international	0.12
say	75.8	3.52	intern	0.11	81.0	0.11	ago	250.69	2,604.7	371.0	public health	0.10
six	70.9	52.85	salt	0.41	63.9	0.41	tribute	249.76	2,593.8	279.7	become	0.07
month	58.9	8.21	health	0.41	42.6	0.41	angels in white	4,679.28	2,592.2	232.0	announced	0.12
million	54.5	16.46	water	0.14	39.7	0.14	algeria	4,670.46	2,586.7	227.2	old traditional	0.02
food	53.8	26.51	public	0.11	32.4	0.11	family members	934.97	2,580.7	220.8	practitioner	0.01
usaheadline	51.4	111.76	bath	0.09	24.1	0.09	2.3	4,657.22	2,578.6	219.3	forget it	0.01
recover	44.7	35.52	organ	0.34	13.0	0.34	ten thousand	4,652.81	2,575.9	218.8	f**k you	0.01
youth	40.9	90.97	isolate	0.22	11.5	0.22	the first branch	4,652.81	2,575.9	218.3	dead man	0.01
Web link URL domains	χ^2	RR	Web link URL domains	RR	χ^2	RR	Web link URL domains	RR	χ^2	RR	Web link URL domains	RR
usaheadlines.today	39.0	84.52	akpraise.com	0.04	11.0	0.04	qq.com	7.15	264.6	207.3	ifeng.com	0.14
foreignpolicy.com	13.0	33.42	facebook.com	0.38	4.4	0.38	xinhuanet.com	3.05	38.4	12.5	infzm.com	0.04
thejournal.ie	10.0	11.14	yabaleftonline.com	0.10	3.2	0.10	mydrivers.com	83.91	36.7	3.1	huanqiu.com	0.24
olodotation.com	8.1	9.83	foxnews.com	0.24	2.9	0.24	gov.cn	65.26	27.3	1.9	chinanews.com	0.44
online.wsj.com	8.1	9.83	japannews.co	0.12	2.7	0.12	cankaoxiaoxi.com	40.40	14.8	1.7	163com	0.63
ynajia.com	5.2	17.69	washingtonpost.com	0.26	2.4	0.26	ithome.com	27.97	8.7	1.7	reuters.com	0.32
nytimes.com	4.6	1.89	ebolafacts.com	0.15	1.7	0.15	sina.cn	2.71	6.4	1.7	sohu.com	0.61
bioportfolio.com	3.3	13.76	time.com	0.39	1.6	0.39	inewthings.com	6.84	6.1	1.4	chinatimes.com	0.18
ndtv.com	3.0	5.90	comeseesumting.com	0.18	1.3	0.18	china.com.cn	21.75	5.7	1.1	zaobao.com	0.47
thetrentonline.com	3.0	5.90	bbc.com	0.18	1.3	0.18	Haochu.com	21.75	5.7	0.8	gzdaily.com	0.24

continued on p. 469

Table 2 (continued). Trending and fading analysis of keywords, Web link URL domains, and hashtags identified in two cross-sectional samples of Twitter tweets and Chinese microblogs (Sina Weibo) on Ebola, August 8-9, 2014, and August 15-16, 2014^a

Hashtags	Twitter ^b						Chinese microblogs (Sina Weibo)					
	Trending			Fading			Trending			Fading		
	χ^2	RR	Hashtags	χ^2	RR	Hashtags	χ^2	RR	Hashtags	χ^2	RR	Hashtags
usaheadlines	41.6	90.07	cdcchat	7.8	0.05	blocking ebola	1,081.8	1,032.64	Sina news prizes share	82.6	0.13	
infographic	3.6	14.66	factsonebola	3.4	0.10	ebola ferocious	6.7	16.54	I have something to say	81.1	0.08	
kenya	3.6	14.66	morelikely	3.0	0.11	information reminded	5.3	13.71	peacekeeping forefront	65.5	0.04	
ebolakenya	3.3	6.28	ebolafact	2.0	0.14	looking at today's topic	3.4	9.92	ebola virus	63.6	0.26	
breaking	1.9	2.09	ebolasyptoms	1.1	0.19	ebola virus prevention	2.5	8.03	death ebola virus	39.7	0.01	
sierraleone	1.7	4.89	anticipate	1.1	0.19	national tourism administration	1.6	6.14	~ NFZM · news	31.1	0.01	
lad	1.7	10.47	yereimito21	0.7	0.23	ebola virus tracking	1.1	5.20	dispelled the rumours	31.1	0.01	
guinea	1.7	10.47	uganda	0.7	0.23	zhengzhou 12320	0.7	4.25	first news	23.3	0.05	
Nano Silver	1.7	10.47	tekmira	0.7	0.23	west three chinese citizens	0.3	3.31	nonsense	22.5	0.02	
Internationaltrade	1.7	10.47	africa	0.6	0.70	snow wright UV germicidal lamp	0.3	3.31	ebola	15.4	0.26	

^aTrending items are those with a high χ^2 value and RR>1. Fading items are those with a high χ^2 value and RR≤1. We expanded the shortened URL Web links contained in the text of the tweet and Chinese microblogs into their original URL Web links and then identified their domains. Hashtags are used in Twitter and Chinese microblogs for a variety of reasons, such as emphasizing the theme of a tweet. For example, #cdcchat is a hashtag used by the Centers for Disease Control and Prevention (CDC) and its followers for their CDC chat activity on Twitter. We translated the Chinese hashtags into English.

^bTerms with the same stem were grouped under one category showing their infinite form. For example, the words "say" and "says" were grouped under "say."

RR = relative risk

Table 3. Content classification of random samples of Ebola-related posts drawn from two cross-sectional samples of Twitter tweets and Chinese microblogs (Sina Weibo) on Ebola, August 8–9, 2014, and August 15–16, 2014

Category	Twitter ^{a,b} Number (percent)		Chinese microblogs (Sina Weibo) ^b Number (percent)	
	Batch 1: August 8–9, 2014	Batch 2: August 15–16, 2014	Batch 1: August 8–9, 2014	Batch 2: August 15–16, 2014
Total posts	245 (100.0)	87 (100.0)	452 (100.0)	207 (100.0)
Outbreak news in general				
News of Ebola outbreak and cases	89 (36.3)	34 (39.1)	167 (36.9)	24 (11.6)
News of travel bans, border blockades, flight route suspensions, sport game bans, and travel advice	1 (0.4)	6 (6.9)	4 (0.9)	7 (3.4)
Assistance to Guinea—Chinese medical team departure for Guinea (news and responses)	NA	NA	3 (0.7)	89 (43.0)
Subtotal	90 (36.7)	40 (46.0)	174 (38.5)	120 (58.0)
Health information in general				
Health education and information	26 (10.6)	10 (11.5)	55 (12.2)	27 (13.0)
Responses to alternative health information	32 (13.1)	10 (11.5)	43 (9.5)	13 (6.3)
Subtotal	58 (23.7)	20 (23.0)	98 (21.7)	40 (19.3)
Alternative health information	6 (2.4)	2 (2.3)	11 (2.4)	3 (1.4)
Other information				
News of a case of someone spreading rumor of “Ebola in Pudong” being detained by police	NA	NA	31 (6.9)	0 (0.0)
Advertisement and entertainment	0 (0.0)	0 (0)	9 (2.0)	2 (1.0)
Social issues	11 (4.5)	0 (0)	15 (3.3)	2 (1.0)
Others	80 (32.7)	25 (28.7)	114 (25.2)	40 (19.3)
Subtotal	91 (37.1)	25 (28.7)	169 (37.4)	44 (21.3)

^aNo Chinese-language tweets were identified in the randomly selected manually coded Twitter samples.

^bOnly posts in English or Chinese were coded. A total of 54 tweets in Batch 1 and 29 tweets in Batch 2 were in a language other than English or Chinese and were not included in the coded sample.

NA = not applicable

restrictions. Moreover, such rumors might have been circulated before or after our sample collections and via other routes or platforms (e.g., word of mouth).

Our findings that 47%–58% of the posts were Ebola-relevant news stories and 19%–24% were health information and responses to alternative health information in both batches are compatible with another analysis of Ebola-related Twitter data from July 24 to August 1, 2014. The four main topics identified in that study were Ebola risk factors, prevention education, disease trends, and prayer for countries in Africa. The authors interpreted their findings as evidence of knowledge gaps about Ebola as relevant information was being provided and sought.²⁹

A study that analyzed 85 of the top 100 YouTube videos on Ebola on December 9, 2014, found that 54 (63.5%) videos had misleading information and 31 (36.5%) videos were useful. However, the study found that the number of views of misleading videos was significantly higher than those of useful videos ($p=0.005$).^{30,31} In contrast, an analysis of 118 videos screened on November 1, 2014, found that 31 (26.3%) YouTube videos were misleading and 87 (73.7%) were useful.³² An analysis of the 100 most widely viewed Ebola-related YouTube videos as of December 9, 2014, found that 36 (36%) reported CDC-described transmission routes of Ebola.³³ Such differences might reflect differences in criteria for content categorization but also the timing of data collection, because social media content evolves continuously.

Two other studies reported an increase in the volume of Ebola-related Twitter traffic during October 2014, when domestic cases were reported in the United States.^{34,35} One exploratory study counted the frequency of keywords associated with different emotional states among Ebola-related tweets.³⁴

We found that key terms used in the WHO's PHEIC announcement were among the top 20 keywords list of the English and Chinese microblogs in Batch 1. This observation might suggest that social media platforms helped disseminate the key WHO messages immediately after the announcement. Our observation that the word "China" occurred more frequently than "Ebola" in Batch 2 of Chinese microblogs was not surprising. Because Chinese government-run media channels emphasized China's anti-Ebola efforts, the term "China" was the most frequently used term in Batch 2 of Chinese microblogs.

Our observation of the increased ratio of Sina Weibo bloggers using hashtags from Batch 1 to Batch 2 might be attributable to hashtags created by some news outlets or the service provider Sina Weibo (e.g.,

a hashtag of a promotional campaign reading "Sina News to share with prizes").

Although many studies attempted to use digital big data (including social media data) to detect outbreaks³⁶ or to estimate or forecast disease incidence,^{37–39} their successful application in public health practice faces technical challenges.^{40–42} Communication surveillance is becoming an important application of social media data in public health surveillance.⁴³ Communication surveillance includes both surveillance of general awareness of certain diseases^{28,44} and monitoring of reactions to public health messages or campaigns.^{2,45}

Implications for practitioners and policy makers

The low proportion of microblogs with alternative health information at the onset of the global response to the 2014–2015 Ebola outbreak mirrors results from studies during the 2009 influenza pandemic, when only about 2% of tweets were seen as misinformation.⁴⁶ We found that most information on social media came from mainstream news agencies, which generally report information from public health agencies. Our findings also indicate a contextual difference between a free and open online platform and a state-regulated online platform in the contents of Ebola-related microblogs. China's Internet market is controlled by the government,⁴⁷ which explains why posts related to misinformation or rumors were not widely observed among Chinese microblogs, whereas misinformation (e.g., saltwater, Nano Silver) was freely distributed on Twitter. For the same reason, topics of discussion on Twitter were more diverse. These findings suggest that a free and open online platform may enable the dissemination of unofficial and unconfirmed information but also pluralistic views. Although censoring allows governments to control rumors and alternative information, it can put the society at risk of a potential government cover-up, as in the initial denial of the 2003 severe acute respiratory syndrome outbreak in China.⁴⁸

Strengths and limitations

The strength of manual coding is the understanding and interpretation of the context and meaning of the content. We compensated for our weakness of having a single manual coder by having a second coder code 10% of the sample, with moderate to substantial reliability. Our random sample (about 4.5 million tweets per day) constituted about 1% of the Twitter universe. Although we did not retrieve all Ebola-related tweets, our sample was representative. Our Twitter data retrieval method allowed us to calculate the incidence rate of tweets that were Ebola-related across a period of time.

Unlike the United States and China, where social media are important channels of disseminating outbreak information,^{4,43,44} the areas most affected by the outbreak are likely to use traditional means of communication to disseminate misinformation. Therefore, the low proportion of misinformation on social media might not reflect the rumors circulating among the public in West Africa, where the epidemic occurred. Additionally, we did not code microblogs written in languages other than English and Chinese. Future studies analyzing tweets in other languages might help confirm the external validity of our findings. Furthermore, Twitter and Sina Weibo may not be representative of the global population, because users are generally young and educated.^{49,50}

CONCLUSION

A small percentage of Ebola-related microblogs contained misinformation; most contained outbreak-related news and scientific health information, echoing the Nigerian success in Ebola health communications via Twitter.⁵ Further studies of health information dissemination on government-censored social media⁴⁷ can serve as a comparison to studies performed on uncensored platforms. A future retrospective longitudinal study of Ebola-related information on microblogs will allow us to investigate how the volume and contents of misinformation changed during this outbreak. Analyzing the sources of misinformation and understanding the process by which rumors are created and circulated through re-tweeting will inform effective public health communication strategies on social media.

The authors thank Sidi Liu and Zhaochong Liu for their technical assistance in the retrieval of Twitter and Sina Weibo data; Tsun Chan and Angela An Chang Cheung for their preliminary manual coding of Weibo posts; Po-Ying Lai for serving as the second coder; and Manoj Gambhir, PhD, and Karen K. Wong, MD, for their helpful discussion.

Isaac Chun-Hai Fung and Zion Tsz Ho Tse receive salary support from the Centers for Disease Control and Prevention (CDC #15IPA1509134 and #16IPA1619505). This project was not funded by CDC and CDC had no role in the study design, data analysis, or writing of this article. This article does not represent the official position of CDC or the U.S. government.

Our protocol of data processing and de-identification was approved by the Human Research Ethics Committee for Non-Clinical Faculties, The University of Hong Kong, and the Georgia Southern University Institutional Review Board (H14167).

Isaac Chun-Hai Fung, King-Wa Fu, and Chung-Hong Chan contributed equally as co-first authors. Thomas Abraham and Zion Tsz Ho Tse contributed equally as co-senior authors.

REFERENCES

- Merino JG. Response to Ebola in the US: misinformation, fear, and new opportunities. *BMJ* 2014;349:g6712.
- Thackeray R, Burton SH, Giraud-Carrier C, Rollins S, Draper CR. Using Twitter for breast cancer prevention: an analysis of breast cancer awareness month. *BMC Cancer* 2013;13:508.
- Harris JK, Mueller NL, Snider D, Haire-Joshu D. Local health department use of Twitter to disseminate diabetes information, United States. *Prev Chronic Dis* 2013;10:E70.
- Fung IC, Wong K. Efficient use of social media during the avian influenza A(H7N9) emergency response. *Western Pac Surveill Response J* 2013;4:1-3.
- Carter M. How Twitter may have helped Nigeria contain Ebola. *BMJ* 2014;349:g6946.
- World Health Organization. Ebola: experimental therapies and rumoured remedies: Ebola situation assessment. 2014 Aug 15 [cited 2014 Aug 22]. Available from: <http://www.who.int/mediacentre/news/ebola/15-august-2014/en>
- Oyeyemi SO, Gabarron E, Wynn R. Ebola, Twitter, and misinformation: a dangerous combination? *BMJ* 2014;349:g6178.
- World Health Organization. Statement on the 1st meeting of the IHR Emergency Committee on the 2014 Ebola outbreak in West Africa. 2014 Aug 8 [cited 2014 Aug 21]. Available from: <http://www.who.int/mediacentre/news/statements/2014/ebola-20140808/en>
- Mendoza M, Poblete B, Castillo C. Twitter under crisis: can we trust what we RT? Proceedings of the First Workshop on Social Media Analytics; 2010 Jul 25–28; Washington.
- Qu Y, Huang C, Zhang P, Zhang J. Microblogging after a major disaster in China: a case study of the 2010 Yushu earthquake. Proceedings of the ACM 2011 Conference on Computer-Supported Cooperative Work; 2011 Mar 19–23; Hangzhou, China.
- Vieweg S, Hughes A, Starbird K, Palen L. Microblogging during two natural hazards events: what Twitter may contribute to situational awareness. Proceedings of the 28th International Conference on Human Factors in Computing Systems; 2010 Apr 10–15; Atlanta.
- World Health Organization. WHO Director-General, Dr. Margaret Chan: #Ebola in West #Africa is a public health emergency of intl concern #alert. Twitter tweet 2014 Aug 8 06:50:00 GMT [cited 2014 Aug 31]. Available from: <https://twitter.com/WHO/status/497635954740715520>
- Twitter Inc. Twitter reports second quarter 2014 results; [news release] 2014 Jul 29 [cited 2014 Aug 31]. Available from: <https://investor.twitterinc.com/releasedetail.cfm?ReleaseID=862505>
- Wehner MR, Chren MM, Shive ML, Resneck JS Jr, Pagoto S, Seidenberg AB, et al. Twitter: an opportunity for public health campaigns. *Lancet* 2014;384:131-2.
- Ayers JW, Althouse BM, Dredze M. Could behavioral medicine lead the Web data revolution? *JAMA* 2014;311:1399-400.
- Weibo Corporation. Weibo reports second quarter 2014 financial results; [press release] 2014 Aug 14 [cited 2014 Aug 31]. Available from: <http://ir.weibo.com/phoenix.zhtml?c=253076&p=irol-newsArticle&ID=1958713&highlight=>
- Salathé M, Freifeld CC, Mekaru SR, Tomasulo AF, Brownstein JS. Influenza A (H7N9) and the importance of digital epidemiology. *N Engl J Med* 2013;369:401-4.
- R Development Core Team. R: Version 3.0.2. Vienna: R Foundation for Statistical Computing; 2013.
- Hornik K, Mair P, Rauch J, Geiger W, Buchta C, Feinerer I. The textcat package for n-gram based text categorization in R. *J Stat Software* 2013;52(6).
- Feinerer I, Hornik K, Meyer D. Text mining infrastructure in R. *J Stat Software* 2008;25:1-54.
- Viterbi AJ. Error bounds for convolutional codes and an asymptotically optimum decoding algorithm. *Information Theory, IEEE Transactions On* 1967;13:260-9.
- Lei L, Qiao G, Dequan Y, Qimin C, editors. Improved Viterbi algorithm-based HMM2 for Chinese words segmentation. International Conference on Computer Science and Electronics Engineering; 2012 Mar 23–25; Hangzhou, Zhejiang China.
- GitHub. Jieba Chinese text segmentation [cited 2015 Nov 21]. Available from: <https://github.com/fxsjy/jieba>
- cURL. Open-source command line tool and library [cited 2015 Nov 21]. Available from: <http://curl.haxx.se>

25. Yang Y, Pedersen JO. A comparative study on feature selection in text categorization. Proceedings from the 14th International Conference on Machine Learning. San Francisco: Morgan Kaufmann Publishers, Inc.; 1997.
26. Manning CD, Raghavan P, Schütze H. Chi² feature selection. 2008 [cited 2015 Nov 22]. Available from: <http://nlp.stanford.edu/IR-book/html/htmledition/feature-selectionchi2-feature-selection-1.html>
27. Viera AJ, Garrett JM. Understanding interobserver agreement: the kappa statistic. *Fam Med* 2005;37:360-3.
28. Fung IC, Hao Y, Cai J, Ying Y, Schaible BJ, Yu CM, et al. Chinese social media reaction to information about 42 notifiable infectious diseases [published erratum appears in *PLoS One* 2015;10:0129525]. *PLoS One* 2015;10:e0126092.
29. Odium M, Yoon S. What can we learn about the Ebola outbreak from tweets? *Am J Infect Control* 2015;43:563-71.
30. Nagpal SJ, Karimianpour A, Mukhija D, Mohan D. Dissemination of "misleading" information on social media during the 2014 Ebola epidemic: an area of concern. *Travel Med Infect Dis* 2015;13:338-9.
31. Nagpal SJ, Karimianpour A, Mukhija D, Mohan D, Brateanu A. YouTube videos as a source of medical information during the Ebola hemorrhagic fever epidemic. *Springerplus* 2015;4:457.
32. Pathak R, Poudel DR, Karmacharya P, Pathak A, Aryal MR, Mahmood M, et al. YouTube as a source of information on Ebola virus disease. *N Am J Med Sci* 2015;7:306-9.
33. Basch CH, Basch CE, Ruggles KV, Hammond R. Coverage of the Ebola virus disease epidemic on YouTube. *Disaster Med Public Health Prep* 2015;9:531-5.
34. Fung IC, Tse ZT, Cheung CN, Miu AS, Fu KW. Ebola and the social media. *Lancet* 2014;384:2207.
35. Rodriguez-Morales AJ, Castaneda-Hernandez DM, McGregor A. What makes people talk about Ebola on social media? A retrospective analysis of Twitter use. *Travel Med Infect Dis* 2015;13:100-1.
36. Brownstein JS, Freifeld CC, Madoff LC. Digital disease detection—harnessing the Web for public health surveillance. *N Engl J Med* 2009;360:2153-5, 2157.
37. Salathe M, Bengtsson L, Bodnar TJ, Brewer DD, Brownstein JS, Buckee C, et al. Digital epidemiology. *PLoS Comput Biol* 2012;8:e1002616.
38. Broniatowski DA, Paul MJ, Dredze M. National and local influenza surveillance through Twitter: an analysis of the 2012–2013 influenza epidemic. *PLoS One* 2013;8:e83672.
39. Ginsberg J, Mohebbi MH, Patel RS, Brammer L, Smolinski MS, Brilliant L. Detecting influenza epidemics using search engine query data. *Nature* 2009;457:1012-4.
40. Olson DR, Konty KJ, Paladini M, Viboud C, Simonsen L. Reassessing Google flu trends data for detection of seasonal and pandemic influenza: a comparative epidemiological study at three geographic scales. *PLoS Comput Biol* 2013;9:e1003256.
41. Santillana M, Zhang DW, Althouse BM, Ayers JW. What can digital disease detection learn from (an external revision to) Google Flu trends? *Am J Prev Med* 2014;47:341-7.
42. Lazer D, Kennedy R, King G, Vespignani A. Big data. The parable of Google Flu: traps in big data analysis. *Science* 2014;343:1203-5.
43. Fung IC, Tse ZT, Fu KW. The use of social media in public health surveillance. *Western Pac Surveill Response J* 2015;6:3-6.
44. Fung IC, Fu KW, Ying Y, Schaible B, Hao Y, Chan CH, et al. Chinese social media reaction to the MERS-CoV and avian influenza A (H7N9) outbreaks. *Infect Dis Poverty* 2013;2:31.
45. Fung IC, Cai J, Hao Y, Ying Y, Chan BS, Tse ZT, et al. Global Handwashing Day 2012: a qualitative content analysis of Chinese social media reaction to a health promotion event. *Western Pac Surveill Response J* 2015;6:34-42.
46. Chew C, Eysenbach G. Pandemics in the age of Twitter: content analysis of Tweets during the 2009 H1N1 outbreak. *PLoS One* 2010;5:e14118.
47. Fu KW, Chan CH, Chau M. Assessing censorship on microblogs in China: discriminatory keyword analysis and the real-name registration policy. *Internet Computing, IEEE* 2013;17:42-50.
48. Abraham T. Twenty-first century plague: the story of SARS. Baltimore: Johns Hopkins University Press; 2007.
49. Fu KW, Chau M. Reality check for the Chinese microblog space: a random sampling approach. *PLoS One* 2013;8:e58356.
50. Pew Research Internet Project. Social networking fact sheet [cited 2014 Jul 16]. Available from: <http://www.pewinternet.org/fact-sheets/social-networking-fact-sheet>