Twitter Conversations and English News Media Reports on Poliomyelitis in Five Different Countries, January 2014 to April 2015

Braydon J Schaible, MPH[1]; Kassandra R Snook, MPH[1]; Jingjing Yin, PhD[1]; Ashley M Jackson, MPH[1]; Jennifer O Ahweyevu, MPH[1]; Muhling Chong, MPH[1]; Zion Tsz Ho Tse, PhD[2]; Hai Liang, PhD[3,4]; King-Wa Fu, PhD[3] ; Isaac Chun-Hai Fung, PhD[1]

ABSTRACT

Introduction: Twitter and media coverage on poliomyelitis help maintain global support for its eradication.

Objective: To test our hypothesis that themes of polio-related tweets and media articles would differ by location of interest (hashtag of country name mentioned in the tweet; country name mentioned in media articles) but would be similar to each other (tweets and media articles) for each location of interest.

Methods: We retrospectively examined a 40% random sample of Twitter data containing the hashtag #polio from January 1, 2014, to April 30, 2015 (N = 79,333), from which we extracted 5 subcorpora each with a co-occurring hashtag #India (n = 5027), #Iraq (n = 1238), #Nigeria (n = 1364), #Pakistan (n = 11,427), and #Syria (n = 2952). We also retrieved and categorized 73 polio-related English-language news stories from within the same timeframe. We assessed the association between polio-related English news themes and the Twitter content. Descriptive analyses and unsupervised machine learning (latent Dirichlet allocation modeling) were conducted on the 5 Twitter subcorpora.

Results: The results of the latent Dirichlet allocation modeling on the specific subcorpora with country co-occurring hashtags showed significant differences between the 5 countries in terms of content. English mass media content focused largely on violence/conflicts and cases of polio, whereas social media focused on eradication and vaccination efforts along with celebrations.

Discussion: Contrary to our hypothesis, our evidence suggests Twitter content differs significantly from English mass media content. Evidence from our study helps inform media monitoring and communications surveillance during global public health crises, such as infectious disease outbreaks, as well as reactions to health promotion campaigns.

INTRODUCTION

Social media has become a vital tool for public health communication globally, given the penetration of the Internet and mobile phones across the world.[1] In the event of a large-scale health threat, emergency risk communication is crucial in any public health response.[2] In addition to information dissemination, public health professionals perform communication surveillance and media monitoring to assess the communication environment in which they interact.[3] Social media can be used in communication surveillance to assess the public's sentiments regarding public health issues and reactions to major public health events.[4] For example, Twitter data were retrieved and monitored in real time during the Ebola and Zika outbreaks, and analysis revealed how Twitter users reacted to news when cases of emerging infectious diseases were imported into their country and when there were local cases.[5-7]

Poliomyelitis (polio) is a vaccine-preventable disease. Thanks to the Global Polio Eradication Initiative, polio (wild poliovirus type 1) remained endemic in only 3 countries (Afghanistan, Nigeria, and Pakistan) as of 2017.[8] The role of communications remains central to the mission and implementation of the Global Polio Eradication Initiative.[9] Prior research identified peaks in polio-related Twitter data corresponding to dates of important news articles such as the Taliban banning the polio vaccine, suggesting Twitter could amplify polio-related news dissemination; and that anti-vaccine sentiments tended to last longer than provaccine sentiments.[10] Additionally, prior research found that television news directly influenced the number of Ebola-related Internet searches and tweets.[11] We conducted this research to obtain a better understanding of Twitter conversations about polio and what inspires them, in the hope it will enable public health professionals to better communicate about polio vaccination efforts and to solicit greater public support for polio eradication efforts.

The purpose of this study was to retrospectively examine a sample of Twitter data containing the hashtag #polio to explore the most common polio-related topics on Twitter from January 1, 2014, to April 30, 2015. Additionally, we examined if any associations existed between news articles from 6 English-language media sources and #polio Twitter content (which was predominantly English). We hypothesized that 1) the distribution of polio-specific Twitter conversation themes among tweets would differ by the location of interest and 2) the distribution of polio-specific themes specific to a location of interest would be similar between Twitter content and English-language media articles. Here, "location of interest" refers to the country that was the topic of Twitter and media conversation and does not refer to the geolocation from which a tweet was posted.

METHODS

A 40% random sample of Twitter data with hashtag #polio from January 1, 2014, to April 30, 2015, was purchased from Gnip Inc/Twitter, San Francisco, CA.

Author Affiliations

[1] Department of Biostatistics, Epidemiology and Environmental Health Sciences, Jiann-Ping Hsu College of Public Health, Georgia Southern University in Statesboro
[2] School of Electrical Engineering and Computer Engineering, College of Engineering, University of Georgia, Athens
[3] Journalism and Media Studies Centre, University of Hong Kong
[4] School of Journalism and Communication, Chinese University of Hong Kong

Corresponding Author

Isaac Chun-Hai Fung, PhD (cfung@georgiasouthern.edu)

ORIGINAL RESEARCH & CONTRIBUTIONS

Twitter Conversations and English News Media Reports on Poliomyelitis in Five Different Countries, January 2014 to April 2015

Table 1. Descriptive statistics of #polio dataset (40% random sample) and subset with country co-hashtags, January 1, 2014, to April 30, 2015[a]

Variable	#Polio	#India and #Polio	#Iraq and #Polio	#Nigeria and #Polio	#Pakistan and #Polio	#Syria and #Polio
Total no. of tweets	79,333	5027	1238	1364	11,427	2952
Friends	329 (123-880)	293 (115-750.5)	372 (19-1015)	443 (168.75-1268.25)	319 (117-880.5)	427.5 (172-1091.25)
Followers	350 (91-1222)	200 (53-867.5)	333 (90.25-1285.5)	608.5 (154-1691.25)	405 (104-1351)	388 (113-1258.25)
Status updates	3792 (891-14,697)	2089 (433.5-8600.5)	4231 (820.5-15,728)	5006.5 (1264.25-17,621.75)	6341 (1477.5-26,088.5)	4148.5 (913.25-16,144.5)
Favorites	153 (14-1111)	56 (7-502)	142.5 (19-1015)	88 (12-784)	283 (26-1617)	167.5 (18.75-1078.75)

[a] Tweets included in each of the 5 subcorpora are all converted to lower case, which included the hashtag followed by the specific country name (in all lower case). Data other than total no. of tweets are presented as median (interquartile range).

Table 2. Frequency of tweets with mentions (percentage) of selected important users in #polio dataset (40% random sample) and subset with country co-hashtags, January 1, 2014, to April 30, 2015

Variable	#Polio (n = 79,333)	#India and #Polio (n = 5027)	#Iraq and #Polio (n = 1238)	#Nigeria and #Polio (n = 1364)	#Pakistan and #Polio (n = 11,427)	#Syria and #Polio (n = 2952)
@WHO	7056 (8.9)	695 (13.8)	176 (14.2)	37 (2.7)	600 (5.3)	946 (32.0)
@CDCgov or @CDCGlobal	1180 (1.5)	99 (2.0)	0 (0)	118 (8.7)	24 (0.2)	2 (0.1)
@NIH	6 (< 0.1)	0 (0)	0 (0)	0 (0)	2 (< 0.1)	0 (0)
@gatesfoundation or @BillGates	2161 (2.7)	1087 (21.6)	0 (0)	55 (4.0)	68 (0.6)	2 (0.1)
@EndPolioNow	1369 (1.7)	17 (0.3)	0 (0)	21 (1.5)	68 (0.6)	3 (0.1)
@UNICEF or @UNICEFpolio	8930 (11.3)	647 (12.9)	538 (43.5)	21 (1.5)	452 (4.0)	591 (20.0)

A total of 79,333 tweets were included in the dataset, all of which contained "#polio" in the body of each tweet. Variables of interest included keywords, hashtags, tweet frequencies, and user statistics. A descriptive analysis was conducted to provide a summary of the entire sample. The numbers of friends, followers, status updates, and favorites for all Twitter accounts in the sample were summarized by the median and the interquartile range, and the frequency and percentages of mentions of certain governmental health-related entities are also provided. The descriptive analysis also found that the number of #polio tweets with a co-occurring hashtag followed by a country name was quite large. Thus, 5 subcorpora of the data were created by extracting tweets, each with a specific country co-occurring hashtag (#India, #Iraq, #Nigeria, #Pakistan, and #Syria), and a similar descriptive analysis was conducted, stratified by each country co-occurring hashtag (Tables 1 and 2). Additional summary statistics are provided, including the top 10 hashtags, keywords, and mentions, along with the distribution of languages of the tweets, and the number of tweets over time (for the entire sample, as well as stratified by country co-occurring hashtag).

Latent Dirichlet allocation (LDA) modeling was conducted on the 5 subcorpora. LDA is a type of unsupervised machine-learning method.[12] An LDA model is a probabilistic topic model. It assumes that documents are "bags of words" and ignores the order of words in a document. The model assumes that across a corpus of documents, there exists a specified number of underlying topics that are hitherto unknown to the researcher, and a probability for each topic is assigned by the model to each tweet. The LDA models were fit to each subcorpus in this study, varying the number of topics from 2 to 100, and the log-likelihood was then determined for each model corresponding to a specific number of topics. The LDA model with the largest log-likelihood and its corresponding number of topics were then chosen for that specific subcorpus. For each model, the top 20 terms for each topic were extracted, as well as representative tweets. The topics for each model were then manually summarized into themes to obtain a more general idea of the topic, as well as to provide a method for descriptive comparisons between models. The number of tweets for each theme within each country is summarized in Table 3. A χ^2 test of association, using 2000 Monte Carlo simulations to calculate the p value, was then conducted to test for association between country and theme.

English news articles related to polio and published online during the study period were extracted from online databases from 2 newspapers, *The New York Times* and *The Independent*; 2 television news stations, Cable News Network (CNN) and Fox News; and 2 radio news stations, National Public Radio and Public Radio International. Each article pertaining to polio was then manually categorized into 1 of 5 general themes (celebrations/polio-free, current eradication efforts, cases/outbreaks/incidence, violence/conflicts, and other) and tagged with the specific country (or countries) of the article's concern. Coding was conducted by 1 primary coder, and a 10% random sample was then double-coded by a secondary coder. The percentage of agreement for categories was 90%; articles not categorized identically were discussed until agreed on. Articles simply mentioning the word *polio* without direct relevance to the topic at hand were not included. Table 4 provides a summary of the news articles used in this study. A χ^2 test of association, using 2000 Monte Carlo simulations to calculate the

ORIGINAL RESEARCH & CONTRIBUTIONS

Twitter Conversations and English News Media Reports on Poliomyelitis in Five Different Countries, January 2014 to April 2015

p value, was conducted to test for association between the English-language media article themes and the countries of interest.

To test for differences between the percentages of themes for each country between the traditional media articles and the tweets (ie, to test for differences between Tables 3 and 4), we conducted a generalized Mantel-Haenszel test.

For each subcorpus, peaks of the time series in Figure 1 were identified manually and confirmed using a standard signal processing technique,[13] which calculates a moving average (and standard deviation) and identifies "peaks" by determining which points fall on a specific number of standard deviations above the moving average. After the peaks were identified,

English-language news articles published around the time of the individual peaks (a 7-day window occurring within a 3-day period either before or after the Twitter peaks) were determined for each subcorpus to get a better understanding of the effect of social media on the selected English-language media outlets, and vice versa.

All data preprocessing and statistical analyses were conducted in R Version 3.2.3 or higher.[14]

RESULTS
Descriptive Statistics of Twitter Data

There were a total of 79,333 tweets in our corpus of a 40% random sample of #polio tweets. Most of the tweets were written in English (n = 68,262; 93.58%),

with the next 4 most common languages being Spanish (n = 5979, 7.54%), French (n = 1197, 1.51%), Italian (n = 1006, 1.27%), and German (n = 575, 0.72%).

The highest number of #polio tweets posted by 1 individual Twitter user was 1307 tweets during the study period. The median number of followers and status updates per user was 350 (Q1 = 91, Q3 = 1222) and 3792 (Q1 = 891, Q3 = 14,697), respectively (Table 1), whereas 8.9% and 11.3% of tweets included case-insensitive mentions for the World Health Organization (WHO), and the United Nations Children's Fund (UNICEF) or UNICEFpolio, respectively (Table 2). The top 5 countries that appeared in co-occurring hashtags in the

Table 3. Themes of topics from latent Dirichlet allocation (LDA) models for each subcorpus with respective country co-occurring hashtag of #India, #Iraq, #Nigeria, #Pakistan, or #Syria, in a 40% random sample of #polio Twitter data, January 1, 2014, to April 30, 2015[a]

Country (no. of tweets)	Theme	Tweets, no. (%)
India (n = 4926)	Celebrations/polio-free	4115 (83.54)
	Current eradication/vaccination efforts	366 (7.43)
	Cases/outbreaks/incidence	233 (4.73)
	Violence/conflicts with polio efforts	0 (0)
	Other	212 (4.30)
Iraq (n = 1221)	Celebrations/polio-free	0 (0)
	Current eradication/vaccination efforts	566 (46.36)
	Cases/outbreaks/incidence	469 (38.41)
	Violence/conflicts with polio efforts	0 (0)
	Other	186 (15.23)
Nigeria (n = 1328)	Celebrations/polio-free	691 (52.03)
	Current eradication/vaccination efforts	476 (35.84)
	Cases/outbreaks/incidence	93 (7.00)
	Violence/conflicts with polio efforts	0 (0)
	Other	68 (5.12)
Pakistan (n = 11,060)	Celebrations/polio-free	0 (0)
	Current eradication/vaccination efforts	4890 (44.21)
	Cases/outbreaks/incidence	1705 (15.42)
	Violence/conflicts with polio efforts	4056 (36.67)
	Other	409 (3.70)
Syria (n = 2879)	Celebrations/polio-free	339 (11.77)
	Current eradication/vaccination efforts	1096 (38.07)
	Cases/outbreaks/incidence	926 (32.16)
	Violence/conflicts with polio efforts	153 (5.31)
	Other	365 (12.68)

[a] Chi-squared = 18,683, p < 0.001. Total number of tweets represent the total number of nonempty tweets used in the models. A small number of "empty tweets" in each of the 5 subcorpora were removed before LDA modeling (after removing stop words, white space, punctuations, etc).

Table 4. Themes of media articles from 6 selected English-language media sources related to polio and 5 countries of interest (India, Iraq, Nigeria, Pakistan, or Syria), January 1, 2014, to April 30, 2015[a]

Country (no. of articles)	Theme	Articles, no. (%)
India (n = 8)	Celebrations/polio-free	6 (75.00)
	Current eradication/vaccination efforts	0 (0)
	Cases/outbreaks/incidence	0 (0)
	Violence/conflicts with polio efforts	0 (0)
	Other	2 (25.00)
Iraq (n = 11)	Celebrations/polio-free	1 (9.09)
	Current eradication/vaccination efforts	1 (9.09)
	Cases/outbreaks/incidence	2 (18.18)
	Violence/conflicts with polio efforts	7 (63.64)
	Other	0 (0)
Nigeria (n = 9)	Celebrations/polio-free	3 (33.33)
	Current eradication/vaccination efforts	1 (11.11)
	Cases/outbreaks/incidence	3 (33.33)
	Violence/conflicts with polio efforts	1 (11.11)
	Other	1 (11.11)
Pakistan (n = 37)	Celebrations/polio-free	1 (2.70)
	Current eradication/vaccination efforts	2 (5.41)
	Cases/outbreaks/incidence	2 (5.41)
	Violence/conflicts with polio efforts	22 (59.46)
	Other	10 (27.03)
Syria (n = 24)	Celebrations/polio-free	1 (4.17)
	Current eradication/vaccination efforts	3 (12.50)
	Cases/outbreaks/incidence	2 (8.33)
	Violence/conflicts with polio efforts	14 (58.33)
	Other	4 (16.67)

[a] Standardized with content themes of a 40% random sample of #polio Twitter data obtained through latent Dirichlet allocation modeling and manual coding. Chi-squared = 50.941, p < 0.001. See the text for the 6 media sources.

ORIGINAL RESEARCH & CONTRIBUTIONS

Twitter Conversations and English News Media Reports on Poliomyelitis in Five Different Countries, January 2014 to April 2015

#polio corpus were: #India (n = 5027), #Iraq (n = 1238), #Nigeria (n = 1364), #Pakistan (n = 11,427), and #Syria (n = 2952) (Table 1). When the descriptive statistics were broken down by these country co-occurring hashtags, the percentage of tweets containing a mention for WHO was much larger for the Syria subcorpus (32.0%) compared with 5.3% for Pakistan, 13.8% for India, 14.2% for Iraq, and just 2.7% for Nigeria (Table 2).

The top 10 co-occurring hashtags, keywords, and mentions of the #polio dataset are found in Table 5. The top 2 co-occurring hashtags (case-sensitive) were #endpolio (n = 11,303) and #Pakistan (n = 10,612), and not surprisingly, the top 3 keywords in the #polio dataset (when all words transformed into lower case) were "polio" (n = 80,156), "pakistan" (n = 14,675), and "endpolio" (n = 12,551). Case-sensitive mentions of @UNICEF (n = 8682) and @WHO (n = 7059) ranked top 2 in the list of mentions in the #polio dataset. A small number of tweets may have contained more than 1 of the same hashtag or mention, which explains the difference in total mentions/hashtags between Tables 2 and 5.

A time trend (daily incidence) of #polio tweets with each of the country-specific co-occurring hashtags can be found in Figure 1.

Themes of Twitter Contents in Five Subcorpora

The results of the LDA modeling on the specific subcorpora with country co-occurring hashtags showed differences between the 5 countries in terms of themes. Table 3 shows the percentage of tweets for each categorized theme from the individual LDA models. In the #India subcorpus, the majority of tweets centered around India achieving a polio-free status, represented with a theme of celebrations/polio-free (n = 4115, 83.54%). The results from the #Iraq subcorpus LDA modeling were much different, indicating that most of the tweets were in regard to current eradication/vaccination efforts (n = 566, 46.36%) and cases/outbreaks/incidence (n = 469, 38.41%). Additionally, most #Nigeria subcorpus tweets were categorized as celebrations/polio-free (n = 691, 52.03%), followed by current eradication/vaccination efforts (n = 476, 35.84%). Regarding

the themes for the #Pakistan subcorpus, 44.21% (n = 4890) of the tweets were categorized as current eradication/vaccination efforts, whereas 36.67% (n = 4056) were categorized as violence/conflicts with polio efforts. Finally, in the #Syria subcorpus, 38.07% (n = 1096) of the tweets were categorized as current eradication/vaccination efforts, and 32.16% (n = 926) as cases/outbreaks/incidence. Examples of #polio tweets for the #India, #Iraq, #Nigeria, #Pakistan, and #Syria subcorpora are found in the Supplemental Figures 1 to 5 (available online at: www.thepermanentejournal.org/files/2019/18-181-Suppl1.pdf), respectively. The χ^2 test of association resulted in a χ^2 statistic of 18,683 (p < 0.001), suggesting a significant association between the coun-tries of interest and the frequency of tweets in the 5 specific themes. In other words, as confirming our first hypothesis, there were statistically significant differences in the percentage of tweets for each theme across the 5 countries of interest.

Analysis of Traditional Media Sources

The accompanying analysis of articles produced by the 6 selected English-language media outlets during the study period can be found in Table 4. In our sample of English-language media, within the timeframe of January 1, 2014, to April 30, 2015, we identified 28 stories in *The New York Times*, 9 in *The Independent*, 1 from CNN, 6 from Fox News, 18 from National Public Radio, and 11 from Public Radio International that discussed polio in 1 or more of the 5 countries of interest. The

results show that 6 of the 8 articles about India were celebrating India's achievement of being polio free (with the breakdown of articles by news source as follows): *The New York Times*, 1/1; *The Independent*, 1/2; CNN, 0/1; Fox News, 0/0; National Public Radio, 2/2; Public Radio International, 2/2). In contrast, most traditional news articles on polio in Iraq, Pakistan, and Syria were focused on violence/conflicts with polio efforts (63.64%, 59.46%, and 58.33%, respectively). In the case of Nigeria, one-third of the traditional news articles were categorized as celebrations/polio-free, and another one-third were in regard to new cases, outbreaks, or the incidence of polio (Table 4). Details of these English-language media articles can be found in the Supplemental Tables 1 to 5 (available online at: www.thepermanentejournal.org/files/2019/18-181-Suppl2.pdf). The χ^2 test of association resulted in a χ^2 statistic of 50.941 (p < 0.001), suggesting that there is a significant association between the themes and countries.

Contrary to our second hypothesis, the generalized Mantel-Haenszel test resulted in a statistic of 17,462 (p < 0.001), suggesting that the frequencies of themes for each country are significantly different between the tweets and the English-language media articles.

Last, we found that 23 English-language media news reports from the 6 selected media sources occurred within a 7-day window of Twitter peaks. Although no concrete statistical associations can be made, about half of traditional news articles (12 of 23)

Table 5. Top 10 most frequent co-occurring hashtags, keywords, and mentions for entire #polio dataset (40% random sample), January 1, 2014, to April 30, 2015[a]

Hashtag	Number	Keyword	Number	Mention	Number
#endpolio	11,303	polio	80,156	@UNICEF	8682
#Pakistan	10,612	pakistan	14,675	@WHO	7059
#vaccineswork	5039	endpolio	12,551	@AseefaBZ	3956
#India	4872	unicef	9248	@rotary	1816
#Syria	2705	vaccine	8295	@UNICEFIndia	1583
#vaccine	1825	india	7288	@EndPolioNow	1374
#ppnews	1633	amp	7285	@BillGates	1359
#WorldPolioDay	1441	cases	7262	@etribune	1204
#Nigeria	1322	via	6572	@UNICEFmena	1002
#Karachi	1286	children	6454	@CDCGlobal	976

[a] Top hashtags and mentions are case sensitive, whereas keywords were transformed to lower case before frequencies were tabulated. In addition, a small number of tweets may have contained more than one of the same hashtag or mention, which explains the difference in total mentions/hashtags between Tables 2 and 5.

ORIGINAL RESEARCH & CONTRIBUTIONS

Twitter Conversations and English News Media Reports on Poliomyelitis in Five Different Countries, January 2014 to April 2015

were published within 3 days *after* an identified Twitter peak of the same subcorpus. Additionally, 7 of the 12 news articles occurring within 3 days after a Twitter peak were identified in the #Pakistan subcorpus.

DISCUSSION

As hypothesized, polio-specific Twitter conversation themes differed by the location of interest. Most of the topics overall were related to the eradication efforts and status, with more specific topics being eradication strategies or vaccinations. For example, a large percentage (83.54%) of the #India subcorpus (Tweets containing both #India and #polio) were affirming that India was now polio free, a declaration made by the WHO in 2014.[15] Similarly, in the #Nigeria co-occurring hashtag subcorpus, 52.03% of the tweets (691/1328) were categorized as "Nigeria polio free soon." Twitter users had expectations that polio would soon be eradicated from Nigeria. However, in the other 3 countries, ongoing armed conflicts had hindered the progress of polio eradication efforts. These were reflected in the Twitter conversation in the subcorpora of #Iraq, #Syria, and #Pakistan.

A large outbreak of polio began in Syria in 2013 and spread to Iraq in 2014.[16,17] The themes of the tweets with co-occurring hashtag #Syria reflected the international concern about that outbreak, with themes highlighting the ongoing polio outbreak (926/2879; 32.16%), polio eradication/vaccination efforts (1096/2879; 38.07%), as well as violence that caused conflicts to those efforts (153/2879; 5.31%). When Syria was declared polio free, it was celebrated on Twitter (339/2879; 11.77%). Likewise, for the subcorpus with #Iraq, nearly 85% of the tweets were about eradication/vaccination efforts (566/1221; 46.36%) and cases/outbreaks/incidence (469/1221; 38.41%).

Similarly, political instability in Pakistan had made polio eradication very difficult. Since the death of Osama Bin Laden, radical Islamist militants had increased their hostility and violence against community health workers who provided polio vaccinations.[18,19] For example, 7 in 10 tweets with co-occurring hashtag #Pakistan were pertinent to violence/conflicts with polio efforts (4056/11,060; 36.67%) or current eradication/vaccination efforts (4890/11,060; 44.21%).

Reviewing the topics for each of the 5 countries (India, Iraq, Nigeria, Pakistan, and Syria), one can clearly see some similarities and differences in what is important in creating or maintaining progress toward the eradication of polio. For instance, because vaccination is the key intervention with which polio can be eradicated globally, there is no surprise that each country co-occurring hashtag had a theme of vaccination. It is interesting to note that there were a large number of mentions for either @UNICEF or @UNICEFpolio, and a large portion of tweets in the subcorpora of the co-occurring hashtags #Syria and #Iraq contained such mentions. Because polio eradication efforts rely on international donors who, in turn, rely on global public support, we speculate that communicating information regarding the plight of children in conflict-ridden countries such as Syria and Iraq would help maintain the attention of global Twitter users and solicit their support for the polio eradication efforts in those countries.[16,20] Additionally, these mentions may help lead readers to important child-related polio information. Last, there were themes of violence and conflict present among the tweets in the subcorpora of co-occurring hashtags #Pakistan
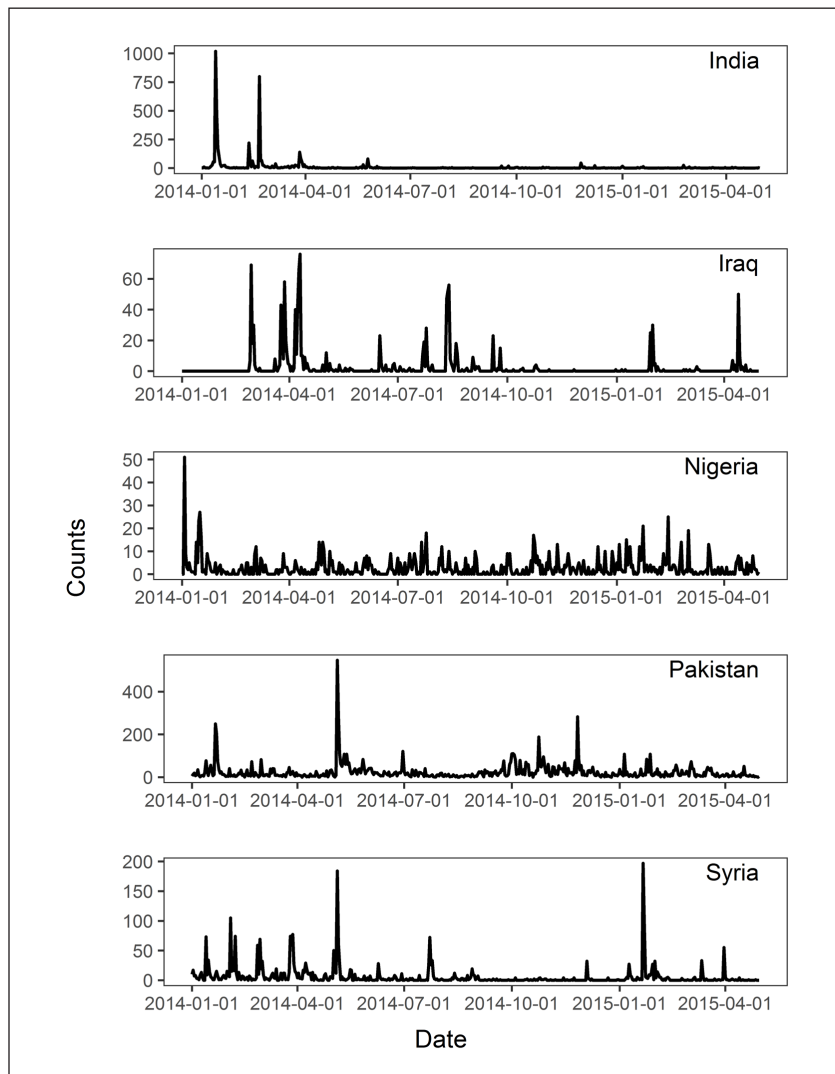

Figure 1. Daily count of tweets over time for 5 subcorpora of a 40% random sample of #polio tweets with co-occurring hashtags, #India, #Iraq, #Nigeria, #Pakistan, and #Syria, from January 1, 2014, to April 30, 2015.

ORIGINAL RESEARCH & CONTRIBUTIONS

Twitter Conversations and English News Media Reports on Poliomyelitis in Five Different Countries, January 2014 to April 2015

and #Syria, signifying the negative impact of certain radical politicoreligious groups on the process of polio eradication and public health in general.[18,19] Therefore, it is important to consider how other factors, such as politics, play a role in polio eradication and in global health security at large. Building trust with public health institutions is of great importance, especially when communities are skeptical toward the international public health community because of religious or political concerns.[2,18]

When examining our sample of articles from the traditional media, there were more articles pertaining to polio and Syria (n = 24) or Pakistan (n = 37) than each of the other 3 countries (India, n = 8; Iraq, n = 11; Nigeria, n = 9). Furthermore, most of these articles focused on violence/conflicts with polio efforts for Syria (14/24, 58.33%) and for Pakistan (22/37, 59.46%).

Contrary to our hypothesis, the results of the generalized Mantel-Haenszel test support the idea that Twitter content differs significantly from English-language media content. Themes with positive undertones such as celebrations/polio-free and current eradication/vaccination efforts were more common in our Twitter sample than in our English-language media article sample for each country of interest. In contrast, English-language media tended to focus on violence, conflicts, and new cases or outbreaks of polio for 4 of the 5 countries under study (except India). Negative topics revolving around polio eradication efforts reported by English-language media may hinder public support of these efforts. Therefore, social media may be a better source for soliciting global health support compared with English-language media.

Prior research has highlighted the role of Twitter as a means to disseminate information pertinent to public health-related topics such as vaccinations and antibiotic use.[21,22] In the past, real-time monitoring of the communication environment of traditional media and social media has provided public health professionals with insights into the concerns of the general public pertinent to infectious disease outbreaks and how public health agencies may respond to their concerns, such as during the 2009 H1N1 influenza pandemic[3] and the 2014 to 2016 Ebola epidemic in West Africa.[23] Analyses of Twitter data help

provide snapshots of the online conversations pertinent to global health at large and on specific diseases in particular[5-7,24]; these may help public health communicators to identify how different linguocultural communities react to ongoing outbreaks, which may differ from the overall global reaction.[25] Meanwhile, a 2015 study on Ebola-related news videos and Twitter data identified Twitter spikes as reactions to news videos released by news organizations.[11] Contrary to this, the current study found that 12 of 23 polio-related English news articles occurred *after* a Twitter peak. However, the differences in content between Twitter and English-language media articles render it difficult to conclude that one affects the other, or vice versa. Regardless, our retrospective analysis of #polio tweets and polio-related news stories published by selected English-language media organizations could provide additional insights on the English-language communication environment pertinent to the polio eradication effort that had run for years with occasional outbreaks in conflict-ridden countries. Our study adds to the literature how themes of Twitter and English-language media coverage may differ when polio is reported in a given country of concern.

Our study is relevant to the Global Health Security Agenda (GHSA). Envisioning a better world that is safe from the threats of infectious disease outbreaks, the GHSA was developed to prevent, detect, and respond to public health emergencies of international concern through international partnership.[26] Effective 2-way communication is vital to harness global support to the GHSA so it can promote health and prevent disease globally. Social media data analysis, as demonstrated by this study and others, can assist public health professionals to better monitor the communication environment and thus better communicate health information to the public.

Our study is not without limitations. Individuals who are at the highest risk of contracting polio are unlikely to be on Twitter. Instead of being a disease outbreak surveillance tool, the #polio Twitter corpus reflected the global Twitter user community's attention to the polio eradication campaign and the challenges associated with it. We want to emphasize that the intention of this study was *not* comparing

polio-related Twitter content tweeted by users *in* these 5 countries. Instead, we compared Twitter content tweeted by users globally *about* these 5 countries. Although our corpus did not contain all tweets with the hashtag #polio, our sample is a 40% random sample without any geographic restriction. The large number of tweets (N = 79,333) reflected a sizable engagement with Twitter users from all over the world. However, we must note that given our data retrieval criteria (hashtag #polio), tweets written in other languages (eg, Arabic and Urdu) were not picked up in our dataset of tweets, which were predominately written in English (93.58%). Also, our media sample was limited to 6 different English-language sources, which were a convenience sample of media organizations. Thus, our results provide only a snapshot of the English-language media environment and may not be representative of the entire English-language media environment. Also, non-English-language local reactions to polio and eradication efforts are beyond the scope of this study. Nevertheless, a better understanding of the English-language international communication environment and the interaction between English-language media and social media remains important. Major technical support for the Global Polio Eradication Initiative continues to come from the US Centers for Disease Control and Prevention.[8,27] The American public support for the Centers for Disease Control and Prevention's participation to eradicate polio globally remains crucial. In this instance, we categorized the English-language news reports during the period under study, and we identified the news reports that were temporally associated with spikes of elevated Twitter activity about polio in 5 countries whose respective co-occurring hashtags appeared to be the top 5 country-specific hashtags within the #polio Twitter corpus.

CONCLUSION

This retrospective analysis of #polio tweets and polio-related English media stories during a 16-month period from January 1, 2014, to April 30, 2015, highlighted the themes of polio-related Twitter conversations as well as news coverage by 6 selected major English-language news media organizations. We identified 23

ORIGINAL RESEARCH & CONTRIBUTIONS

Twitter Conversations and English News Media Reports on Poliomyelitis in Five Different Countries, January 2014 to April 2015

English-language media reports that were within the 7-day window of the spikes in the daily incidence of #polio tweets. Our study provided public health professionals with quantitative data that described the communication environment that might, in turn, affect public perception, and thus support, of the global polio eradication campaign. Because differences in polio-specific themes between social and mass media were identified, further research may elucidate the relationships between the 2 forms of communication. Social media data can assist public health professionals in pinpointing the general concerns or needs of the public during infectious disease events to create location-specific campaigns.[25] Similarly, social media could serve as a positive environment where support can be harnessed for these campaigns. ❖

Disclosure Statement
The author(s) have no conflicts of interest to disclose.

Authors' Contributions
Braydon J Schaible, MPH, and Kassandra R Snook, MPH, are primary coauthors. Braydon J Schaible, MPH; Kassandra R Snook, MPH; Jingjing Yin, PhD; and Isaac Chun-Hai Fung, PhD, conceived the research idea and began the project when Braydon J Schaible, MPH, and Kassandra R Snook, MPH were graduate students of the Jiann-Ping Hsu College of Public Health, Georgia Southern University, Statesboro, GA. Braydon J Schaible, MPH, performed the statistical analysis of Twitter data under the supervision of Jingjing Yin, PhD, and Isaac Chun-Hai Fung, PhD. Kassandra R Snook, MPH, performed the analysis of traditional media articles (with Braydon J Schaible, MPH, as the second coder) under the supervision of Isaac Chun-Hai Fung, PhD. Jennifer O Ahweyevu, MPH, and Muhling Chong, MPH, performed the initial round of manual thematic coding of topics of Twitter data identified by unsupervised machine-learning, under the supervision of Ashley M Jackson, MPH, who was in turn supervised by Isaac Chun-Hai Fung, PhD. Braydon J Schaible, MPH, and Kassandra R Snook, MPH, wrote the early drafts of the manuscript. Jingjing Yin, PhD; Zion Tsz Ho Tse, PhD; Hai Liang, PhD; and King-Wa Fu, PhD, provided intellectual input to this project and edited the manuscript. Isaac Chun-Hai Fung, PhD, made major edits to the manuscript and serves as the senior and corresponding author.

References
1. Chou WY, Hunt YM, Beckjord EB, Moser RP, Hesse BW. Social media use in the United States: Implications for health communication. J Med Internet Res 2009 Oct-Dec;11(4). DOI: https://doi.org/10.2196/jmir.1249.
2. Sell TK. When the next disease strikes: How to communicate (and how not to). Health Secur 2017 Jan-Feb;15(1):28-30. DOI: https://doi.org/10.1089/hs.2016.0100.
3. Prue CE, Lackey C, Swenarski L, Gantt JM. Communication monitoring: Shaping CDC's emergency risk communication efforts. J Health Commun 2003;8(Suppl 1):35-49; discussion 148-51. DOI: https://doi.org/10.1080/713851975.
4. Kostkova P, Brewer H, de Lusignan S, et al. Who owns the data? Open data for healthcare. Front Public Health 2016;4:7. DOI: https://doi.org/10.3389/fpubh.2016.00007.
5. Fu KW, Liang H, Saroha N, Tse ZTH, Ip P, Fung ICH. How people react to Zika virus outbreaks on Twitter? A computational content analysis. Am J Infect Control 2016 Dec 1;44(12):1700-2. DOI: https://doi.org/10.1016/j.ajic.2016.04.253.
6. Fung IC, Fu KW, Chan CH, et al. Social media's initial reaction to information and misinformation on Ebola, August 2014: Facts and rumors. Public Health Rep 2016 May-Jun;131(3):461-73. DOI: https://doi.org/10.1177/003335491613100312.
7. Fung IC, Tse ZT, Cheung CN, Miu AS, Fu KW. Ebola and the social media. Lancet 2014 Dec 20;384(9961):2207. DOI: https://doi.org/10.1016/S0140-6736(14)62418-1.
8. Patel M, Cochi S. Addressing the challenges and opportunities of the polio endgame: Lessons for the future. J Infect Dis 2017 Jul 1;216(suppl 1):S1-8. DOI: https://doi.org/10.1093/infdis/jix117.
9. Menning L, Garg G, Pokharel D, et al. Communications, immunization, and polio vaccines: Lessons from a global perspective on generating political will, informing decision-making and planning, and engaging local support. J Infect Dis 2017 Jul 1;216(suppl 1):S24-32. DOI: https://doi.org/10.1093/infdis/jix059.
10. Bahk CY, Cumming M, Paushter L, Madoff LC, Thomson A, Brownstein JS. Publicly available online tool facilitates real-time monitoring of vaccine conversations and sentiments. Health Aff (Millwood) 2016 Feb;35(2):341-7. DOI: https://doi.org/10.1377/hlthaff.2015.1092.
11. Towers S, Afzal S, Bernal G, et al. Mass media and the contagion of fear: The case of Ebola in America. PloS One 2015 Jun 11;10(6). DOI: https://doi.org/10.1371/journal.pone.0129179.
12. Blei DM. Probabilistic topic models. Commun ACM 2012 Apr;55(4):77-84. DOI: https://doi.org/10.1145/2133806.2133826.
13. Jean-Paul. Peak signal detection in realtime timeseries data [Msg 104]. Message posted to Stackoverflow [Internet]. 2014 [cited 2017 Dec 29]. Available from: https://stackoverflow.com/questions/22583391/peak-signal-detection-in-realtime-timeseries-data.
14. R Core Team. R: A language and environment for statistical computing [Internet]. Vienna, Austria: R Foundation for Statistical Computing; 2015 [cited 2018 Feb 14]. Available from: www.R-project.org.
15. Bahl S, Kumar R, Menabde N, et al. Polio-free certification and lessons learned—South-East Asia region, March 2014. MMWR Morb Mortal Wkly Rep 2014 Oct 24;63(42):941-6.
16. Mbaeyi C, Ryan MJ, Smith P, et al. Response to a large polio outbreak in a setting of conflict—Middle East, 2013-2015. MMWR Morb Mortal Wkly Rep 2017 Mar 3;66(8):227-31. DOI: https://doi.org/10.15585/mmwr.mm6608a6.
17. World Health Organization Regional Office for the Eastern Mediterranean. Polio Eradication Initiative: Iraq [Internet]. 2017 [cited 2018 Feb 14]. Available from: www.emro.who.int/polio/countries/iraq.html.
18. Gostin LO. Global polio eradication: Espionage, disinformation, and the politics of vaccination. Milbank Q 2014 Sep;92(3):413-7. DOI: https://doi.org/10.1111/1468-0009.12065.
19. Kennedy J, McKee M, King L. Islamist insurgency and the war against polio: A cross-national analysis of the political determinants of polio. Global Health 2015 Sep 30;11:40. DOI: https://doi.org/10.1186/s12992-015-0123-y.
20. van Berlaer G, Elsafti AM, Al Safadi M, et al. Diagnoses, infections and injuries in Northern Syrian children during the civil war: A cross-sectional study. PloS One 2017 Sep 8;12(9). DOI: https://doi.org/10.1371/journal.pone.0182770.
21. Love B, Himelboim I, Holton A, Stewart K. Twitter as a source of vaccination information: Content drivers and what they are saying. Am J Infect Control 2013 Jun;41(6):568-70. DOI: https://doi.org/10.1016/j.ajic.2012.10.016.
22. Scanfeld D, Scanfeld V, Larson EL. Dissemination of health information through social networks: Twitter and antibiotics. Am J Infect Control 2010 Apr;38(3):182-8. DOI: https://doi.org/10.1016/j.ajic.2009.11.004.
23. Bedrosian SR, Young CE, Smith LA, et al. Lessons of risk communication and health promotion—West Africa and United States. MMWR Suppl 2016 Jul 8;65(3):68-74.
24. Fung IC, Jackson AM, Ahweyevu JO, et al. #Globalhealth Twitter conversations on #Malaria, #HIV, #TB, #NCDS, and #NTDS: A cross-sectional analysis. Ann Global Health 2017 May-Aug;83(3-4):682-90. DOI: https://doi.org/10.1016/j.aogh.2017.09.006.
25. Fung IC-H, Zeng J, Chan C-H, et al. Twitter and Middle East respiratory syndrome, South Korea, 2015: A multi-lingual study. Infect Dis Health 2018 Mar;23(1):10-16. DOI: https://doi.org/10.1016/j.idh.2017.08.005.
26. Wolicki SB, Nuzzo JB, Blazes DL, Pitts DL, Iskander JK, Tappero JW. Public health surveillance: At the core of the Global Health Security Agenda. Health Secur 2016 May-Jun;14(3):185-8. DOI: https://doi.org/10.1089/hs.2016.0002.
27. Who we are: CDC and the Global Polio Eradication Initiative [Internet]. Atlanta, GA: Centers for Disease Control and Prevention; 2013 [cited 2018 Feb 14]. Available from: www.cdc.gov/polio/who.