Article:
Fung, I.C.-H., Jackson, A.M., Ahweyevu, J.O. et al. (6 more authors) (2017) #Globalhealth Twitter Conversations on #Malaria, #HIV, #TB, #NCDS, and #NTDS: a Cross-Sectional Analysis. Annals of global health. pp. 682-690. ISSN 2214-9996

https://doi.org/10.1016/j.aogh.2017.09.006

Annals of Global Health
VOL. 83, NO. 3–4, 2017

ISSN 2214-9996/$36.00
https://doi.org/10.1016/j.aogh.2017.09.006

ORIGINAL RESEARCH

#Globalhealth Twitter Conversations on #Malaria, #HIV, #TB, #NCDS, and #NTDS: a Cross-Sectional Analysis

Isaac Chun-Hai Fung, PhD[1], Ashley M. Jackson, MPH[1], Jennifer O. Ahweyevu, BS[2], Jordan H. Grizzle, MPH[2], Jingjing Yin, PhD, Zion Tsz Ho Tse, PhD, Hai Liang, PhD, Juliet N. Sekandi, MD, MS, DrPH, King-Wa Fu, PhD
Statesboro, GA; Athens, GA; Hong Kong; Cambridge, MA

Abstract

BACKGROUND Advocates use the hashtag #GlobalHealth on Twitter to draw users' attention to prominent themes on global health, to harness their support, and to advocate for change.

OBJECTIVES We aimed to describe #GlobalHealth tweets pertinent to given major health issues.

METHODS Tweets containing the hashtag #GlobalHealth (N = 157,951) from January 1, 2014, to April 30, 2015, were purchased from GNIP Inc. We extracted 5 subcorpora of tweets, each with 1 of 5 co-occurring disease-specific hashtags (#Malaria, #HIV, #TB, #NCDS, and #NTDS) for further analysis. Unsupervised machine learning was applied to each subcorpus to categorize the tweets by their underlying topics and obtain the representative tweets of each topic. The topics were grouped into 1 of 4 themes (advocacy; epidemiological information; prevention, control, and treatment; societal impact) or miscellaneous. Manual categorization of most frequent users was performed. Time zones of users were analyzed.

FINDINGS In the entire #GlobalHealth corpus (N = 157,951), there were 40,266 unique users, 85,168 retweets, and 13,107 unique co-occurring hashtags. Of the 13,087 tweets across the 5 subcorpora with co-occurring hashtag #malaria (n = 3640), #HIV (n = 3557), #NCDS (noncommunicable diseases; n = 2373), #TB (tuberculosis; n = 1781), and #NTDS (neglected tropical diseases; n = 1736), the most prevalent theme was prevention, control, and treatment (4339, 33.16%), followed by advocacy (3706, 28.32%), epidemiological information (1803, 13.78%), and societal impact (1617, 12.36%). Among the top 10 users who tweeted the highest number of tweets in the #GlobalHealth corpus, 5 were individual professionals, 3 were news

The data were purchased using I.C.H.F.'s startup funds at Georgia Southern University. I.C.H.F. and Z.T.H.T. received salary support from the Centers for Disease Control and Prevention (16IPA1609578 and 16IPA1619505). This paper is not related to their Centers for Disease Control and Prevention (CDC)–funded project. The opinions expressed in this review do not represent the CDC or the United States Government.
All coauthors declare that they do not have any conflicts of interest.
I.C.H.F. conceived the research idea, discussed the research idea with Z.T.H.T., and purchased the data. H.L. helped process the data from JSON format into R format. A.M.J. and J.H.G. performed the preliminary data analysis as a class project in the Epidemiology of Infectious Disease course (EPID 7135, Fall 2016) under the instruction of I.C.H.F. and prepared the first draft of the manuscript as a piece of coursework. I.C.H.F. rewrote the manuscript, and then instructed A.M.J. and J.O.A. to redo the data analysis in Spring 2017. A.M.J. edited the manuscript with the new results. I.C.H.F. double-checked the R codes and the results, performed additional analysis, rewrote the manuscript and wrote the appendix. J.Y., Z.T.H.T., K.W.F., H.L. and J.B.S. contributed intellectual inputs to this project by providing feedback to I.C.H.F. and his students, and edited the manuscript. I.C.H.F. and A.M.J. are co-first authors. J.O.A. and J.H.G. are co-second authors. A.M.J., J.H.G., and J.O.A. are students of I.C.H.F. and J.Y. I.C.H.F. is the senior and corresponding author.
[1]These two authors contributed equally as co-first authors.
[2]These two authors contributed equally as co-second authors.
From the Department of Epidemiology and Environmental Health Sciences, Jiann-Ping Hsu College of Public Health, Georgia Southern University, Statesboro, GA (IC-HF, AMJ, JOA, JHG); Department of Biostatistics, Jiann-Ping Hsu College of Public Health, Georgia Southern University, Statesboro, GA (JY); School of Electrical and Computer Engineering, The University of Georgia, Athens, GA (ZTHT); School of Journalism and Communication, Chinese University of Hong Kong, Hong Kong (HL); Global Health Institute, The University of Georgia, Athens, GA (JNS); Department of Epidemiology and Biostatistics, The University of Georgia, Athens, GA (JNS); Journalism and Media Studies Centre, The University of Hong Kong, Hong Kong (K-WF); and MIT Media Lab, Massachusetts Institute of Technology, Cambridge, MA (K-WF). Address correspondence to I.C.-H.F. (cfung@georgiasouthern.edu).

media, and 2 were organizations advocating for global health. The most common users' time zone was Eastern Time (United States and Canada).

CONCLUSIONS This study highlighted the specific #GlobalHealth Twitter conversations pertinent to malaria, HIV, tuberculosis, noncommunicable diseases, and neglected tropical diseases. These conversations reflect the priorities of advocates, funders, policymakers, and practitioners of global health on these high-burden diseases as they presented their views and information on Twitter to their followers.

KEY WORDS global health, health communication, Internet, machine learning, manual coding, social media, Twitter.

INTRODUCTION

Global health challenges persist in this age of technological advances. According to the Global Burden of Disease 2015 study,[1] of all 55.8 million deaths globally in 2015, noncommunicable diseases (NCDS) resulted in 39.8 million deaths (71.3%). Likewise, human immunodeficiency virus/acquired immunodeficiency syndrome (HIV/AIDS) caused 1.2 million deaths (2.1%); tuberculosis (TB) led to 1.1 million deaths (2.0%); and 843.1 thousand deaths (1.5%) were due to neglected tropical diseases (NTDS) and malaria combined.[1] Generated from pre-existing ideas of public health, tropical medicine, and international health, the concept of "global health" highlights health challenges prevalent in less-developed countries, as well as the existing health inequity within and across national borders.[2] To address such challenges requires global collaborative efforts that go beyond the nation-states. Nongovernmental organizations and private citizens are increasingly involved in the conversation about and the mobilization toward addressing these global health concerns.

As social media use is on the rise globally,[3] public health professionals have tapped into the potential of social media as communication platforms to harness support and to raise awareness of important global health issues, as seen in 2 recent global public health emergencies, Ebola and Zika.[4-6] Twitter is a social media platform where users post a "tweet" of 140 characters or less to those who follow their profiles. The platform has approximately 313 million active users monthly, and 79% of the accounts are outside of the United States.[7] Given its global reach, it will be of interest to shed light on how Twitter users discuss global health. Many users use hashtags (#) to highlight keywords as they engage in Twitter conversations with other users. The hashtag #GlobalHealth highlights conversations focusing on global health. Furthermore, users may use multiple hashtags in a single tweet to highlight interrelated concepts. For example, a user may use hashtags #GlobalHealth and #Malaria to highlight malaria as a global health issue.

Communication monitoring conducted by public health agencies, such as the Centers for Disease Control and Prevention,[8] has recently been extended to cover social media platforms. Health communicators can now "listen" to the communication environment in which they operate, whether it is their routine operating environment or risk communication during emergency responses, such as the 2014-2016 Ebola outbreak.[9] As discussed in a systematic review of Ebola-related social media research,[4] researchers have been working toward the implementation of both qualitative and quantitative methods (including machine learning) to the corpora of the large volume of social media posts. Although social media research pertinent to specific global health concerns, such as Ebola and Zika outbreaks, have been conducted, it is also important to study the Twitter communication environment that is specific to the topic of global health.

The aim of this study was to describe and categorize a cross-sectional Twitter dataset of 16 months (January 2014 to April 2015) with the hashtag #GlobalHealth. We chose the corpus with the specific hashtag #GlobalHealth to focus on tweets that are genuinely discussing the topic of global health (increasing specificity and reducing noise). More specifically, our analysis focused on 5 subcorpora with co-occurring hashtags #Malaria, #HIV, #NCDS, #TB, and #NTDS (the 5 most commonly co-occurring hashtags after #Ebola in our #GlobalHealth dataset). Our study sheds light on how issues of global health were discussed on Twitter, and public health professionals may use this information as a stepping stone toward better communicating global health needs on social media.

METHODS

All tweets containing the hashtag #GlobalHealth (n = 157,951) from January 1, 2014, to April 30, 2015,

Annals of Global Health, VOL. 83, NO. 3–4, 2017
May–August 2017: 682–690

were purchased from GNIP Inc., a subsidiary of Twitter, Inc. We first identified the top co-occurring hashtags in the corpus. Although #Ebola was the top co-occurring hashtag (this was during the 2014-2015 West African Ebola outbreak), we decided to focus our study on other #GlobalHealth conversations because our research goal was not to focus on the ongoing Ebola outbreak (for Ebola-related social media research, please refer to a systematic review).[4] Excluding #Ebola, the top 5 disease-specific co-occurring hashtags were #Malaria (n = 3640), #HIV (n = 3557), #NCDS (n = 2373), #TB (n = 1781), and #NTDS (n = 1736). We then extracted the 5 subcorpora with these co-occurring hashtags for further analysis. Descriptive statistics of the corpus and each subcorpus, including the time zone of the Twitter users, were obtained. For each subcorpus, the 10 users who posted the highest number of tweets therein, as well as the 10 co-occurring hashtags that appeared most frequently, were also identified and reported.

In addition, we applied an unsupervised machine learning method (known as Latent Dirichlet Allocation analysis) to each subcorpus to identify the underlying topics of the contents. Latent Dirichlet Allocation analysis assumed that a certain number of underlying topics (prespecified by the user) existed within a corpus of documents, and such topics consisted of "bags" of words that were more likely to co-occur in the same document of the topic than in documents of other topics.[10] For each subcorpus, we ran Latent Dirichlet Allocation analysis 5 times for each number of topics (from 5 topics to 100 topics, by an interval of 5). The model with the best "goodness of fit" index (measured by the lowest perplexity score) was selected. For each topic in the selected model, the probability of each tweet belonging to that model was calculated. Each tweet would then be assigned to the topic of which the tweet had the highest probability. The representative tweets of each topic were manually selected by the 2 coders (A.M.J. and J.O.A.) who reviewed the tweets from the Latent Dirichlet Allocation model output representative of each topic. The machine-generated topics were further manually classified into 4 predefined thematic categories. These 4 categories were defined by an epidemiologist (the corresponding author): advocacy; epidemiological information; prevention, control, and treatment; and societal impact. *Advocacy* referred to tweet topics that campaign for a cause (eg, more research being devoted to a specific disease). *Epidemiological information* referred to topics delivering information, such as incidence, prevalence, and case-mortality ratio. *Prevention, control, and treatment* referred to topics that mention actions that humans can take to prevent, control, or treat a specific disease. *Societal impact* referred to topics that mention the broader impact of a disease (eg, the economic impact of an epidemic). Any tweet topics that did not fall into the 4 aforementioned categories were categorized as *Miscellaneous*. The 2 coders (A.M.J., J.O.A.) applied this scheme to the topics, independently grouped the topics into themes, and came to a consensus for each subcorpus, reaching 100% agreement at the end.

The time trends of the daily frequency of tweets in the #GlobalHealth corpus and the 5 subcorpora were also manually inspected. The authors (I.C.H.F., A.M.J., and J.O.A.) manually read the tweets of the selected peaks and identified the news or events that triggered the heightened Twitter activity.

The statistical language R, Version 3.3.1 (Vienna, Austria, R Development Core Team) was used to perform all analyses. Further details of the methods are presented in the Online Supplementary Materials.

Ethics Approval. This research was approved by Georgia Southern University's Institution Review Board (H15083) under the B2 exempt category because the social media posts analyzed in this study are considered publically observable behavior.

RESULTS

Among the tweets in the #GlobalHealth corpus from January 1, 2014, through April 30, 2015 (n = 157,951), there were 40,266 unique users, 85,168 retweets, and 13,107 unique co-occurring hashtags (Table 1). For each of the 5 subcorpora with the co-occurring hashtags #Malaria, #HIV, #NCDS, #TB, and #NTDS, the number of tweets, the number of unique users, the number of retweets, and the number of unique co-occurring hashtags are presented in Table 1. The subcorpus with co-occurring hashtag #Malaria had the largest number of tweets (n = 3640), the largest number of unique users (n = 1,737), the largest number of the retweets (n = 2176), and the highest number of co-occurring hashtags (n = 830) (Table 1). Of the 157,951 tweets, 148,564 (94.1%) were labeled as English ("en") by Twitter. Of the 5 subcorpora under study, English language tweets (as labeled by Twitter) ranged from 90.0% in #Malaria to 95.0% in #NCDS (Table 1). Regarding time zone, the most common time zones in the #GlobalHealth corpus were Eastern Time (United States & Canada) (38,260/157,951, 24.2%), followed by Pacific Time (United States & Canada) (10,093/157,951, 6.4%),

Table 1. Descriptive Statistics of Twitter Corpus With Hashtag #GlobalHealth and 5 Subcorpora With Co-occurring Hashtags

	No. of Tweets,* n (%)	No. of Unique Users, n (%)	No. of Retweets, n (%)	No. of Unique Co-occurring Hashtags, n (%)	Tweets in English, n (%)†
Entire corpus	157,951 (100)	40,266 (100)	85,168 (100)	13,107 (100)	148,564 (94.1)
Subcorpora with co-occurring hashtag					
#Malaria	3640 (2.30)	1737 (4.31)	2176 (2.55)	594 (4.53)	3299 (90.6)
#HIV	3558 (2.25)	1676 (4.16)	1778 (2.09)	830 (6.33)	3343 (94.0)
#NCDS	2373 (1.50)	1126 (2.80)	1531 (1.80)	442 (3.37)	2254 (95.0)
#TB	1781 (1.13)	866 (2.15)	991 (1.16)	348 (2.66)	1691 (94.9)
#NTDS	1736 (1.10)	780 (1.94)	1014 (1.19)	298 (2.27)	1628 (93.8)

NCDS, noncommunicable diseases; NTDS, neglected tropical diseases; TB, tuberculosis.
* These numbers include both original tweets and retweets.
† According to the Twitter metadata.

London (7607/157,951, 4.8%), Central Time (United States & Canada) (6744/157,951, 4.3%) and Atlantic Time (Canada) (6246/157,951, 4.0%) (Table 2). Because many places adopt the practice of daylight saving time during the summer, the frequency of tweets with their exact time difference with coordinated universal time (UTC) are presented in Supplementary Table 1 (see Online Supplementary Materials at doi:10.1016/j.aogh.2017.09.006).

The top 10 Twitter users who tweeted the highest number of tweets with the hashtag #GlobalHealth and the respective top 10 users who tweeted the highest number of tweets with the hashtag #GlobalHealth and one of the 5 co-occurring hashtags (#Malaria, #HIV, #NCDS, #TB, and #NTDS) were divided into 4 categories: (1) individual professionals (including both public health professionals and journalists); (2) news media organizations; (3) organizations, both governmental and nongovernmental (excluding media); and (4) miscellaneous (anyone who did not fall into the aforementioned categories) (Fig. 1; see Supplementary Table 2 for details; it can be found in the Online Supplementary Materials at doi:10.1016/j.aogh.2017.09.006). For example, the top Twitter user in the entire #GlobalHealth corpus was an epidemiologist with a professional focus on global health. Another example was a journalist who wrote for Citizen News Service, an India-based organization of citizen journalists who report on health and science via print and online media.[11] This journalist ranked second in the entire corpus for #GlobalHealth and was among the top 10 in 4 of the 5 subcorpora under study. Citizen News Service, as an example of a news media organization, was among the top 10 in the #HIV subcorpus. An example of organizations was the Deutsche Stiftung Weltbevoelkerung (DSW), or the German Foundation for World Population, a Germany-based international organization that addresses sexual and reproductive health issues. DSW ran a global health–related campaign during the period.[12] DSW ranked top 10 in 4 of the 5 subcorpora (see Supplementary Table 2). It is important to note that in the entire corpus of #GlobalHealth, no single user's posts made up to 1% of the corpus. In the 5 subcorpora, the tweets of the 10th user in the respective lists were around 1% of the respective corpus. In other words, each of the remaining users' shares of the corpus was very small (around 1% or less). Yet there is a long list of such users who contributed few tweets to the corpus, because the top 10 users' share was 6.27% for the #GlobalHealth corpus, and 16.84%, 18.52%, 16.81%, 24.65%, and 21.54% for the #Malaria, #HIV, #NCDS, #TB, and #NTDS subcorpora, respectively.

Supplementary Table 3, which can be found in the Online Supplementary Materials at doi:10.1016/j.aogh.2017.09.006, highlights the 10 most common co-occurring hashtags for each subcorpus. The data represent a minority of tweets that had 3 or more hashtags in a single tweet—that is, #GlobalHealth AND (#Malaria OR #HIV OR #NCDS OR #TB OR #NTDS) AND a third hashtag. These tweets provided the readers with some understanding of how ideas and issues overlap. For example, there were 223 tweets with a combination of 3 hashtags: #GlobalHealth AND #Malaria AND #HIV, highlighting the issue of coinfection of malaria and HIV as an important issue in global health.

Topics identified in the selected Latent Dirichlet Allocation model (with the lowest perplexity score) that was applied to each disease-specific subcorpus were manually categorized into 5 categories (4 themes

Table 2. The 10 Most Common Time Zones of Users Who Posted Tweets in the #GlobalHealth Corpus and 5 Subcorpora

	#GlobalHealth	Frequency (%)	#Malaria and #GlobalHealth	Frequency (%)	#HIV and #GlobalHealth	Frequency (%)	#NCDS and #GlobalHealth	Frequency (%)	#TB and #GlobalHealth	Frequency (%)	#NTDS and #GlobalHealth	Frequency (%)
	Total	157,951 (100)	Total	3640 (100)	Total	3558 (100)	Total	2373 (100)	Total	1781 (100)	Total	1736 (100)
	No data	43,258 (27.4)	No data	1060 (29.1)	No data	992 (27.9)	No data	754 (31.8)	No data	573 (32.2)	No data	389 (22.4)
1	Eastern Time (US & Canada)	38,260 (24.2)	Eastern Time (US & Canada)	721 (19.8)	Eastern Time (US & Canada)	910 (25.6)	Eastern Time (US & Canada)	369 (15.5)	Eastern Time (US & Canada)	295 (16.6)	Eastern Time (US & Canada)	239 (13.8)
2	Pacific Time (US & Canada)	10,093 (6.4)	London	187 (5.1)	Pacific Time (US & Canada)	205 (5.8)	London	211 (8.9)	New Delhi	126 (7.1)	London	231 (13.3)
3	London	7607 (4.8)	Pacific Time (US & Canada)	186 (5.1)	Central Time (US & Canada)	146 (4.1)	Atlantic Time (Canada)	103 (4.3)	Mumbai	82 (4.6)	Central Time (US & Canada)	121 (7.0)
4	Central Time (US & Canada)	6744 (4.3)	Central Time (US & Canada)	123 (3.4)	Atlantic Time (Canada)	141 (4.0)	Amsterdam	68 (2.9)	London	78 (4.4)	Amsterdam	96 (5.5)
5	Atlantic Time (Canada)	6246 (4.0)	Paris	121 (3.3)	London	131 (3.7)	Pacific Time (US & Canada)	62 (2.6)	Pacific Time (US & Canada)	62 (3.5)	Atlantic Time (Canada)	77 (4.4)
6	Quito	4570 (2.9)	Bangkok	118 (3.2)	Quito	122 (3.4)	Copenhagen	59 (2.5)	Central Time (US & Canada)	56 (3.1)	Pacific Time (US & Canada)	60 (3.5)
7	Amsterdam	4293 (2.7)	Amsterdam	115 (3.2)	New Delhi	108 (3.0)	Bern	57 (2.4)	Atlantic Time (Canada)	47 (2.6)	Brussels	56 (3.2)
8	Athens	2735 (1.7)	New Delhi	101 (2.8)	Amsterdam	84 (2.4)	Melbourne	57 (2.4)	Brussels	42 (2.4)	Bern	54 (3.1)
9	New Delhi	2520 (1.6)	Atlantic Time (Canada)	92 (2.5)	Athens	71 (2.0)	Central Time (US & Canada)	54 (2.3)	Chennai	38 (2.1)	Quito	43 (2.5)
10	Bern	2421 (1.5)	Athens	86 (2.4)	Brussels	62 (1.7)	Quito	50 (2.1)	Amsterdam	36 (2.0)	Athens	32 (1.8)

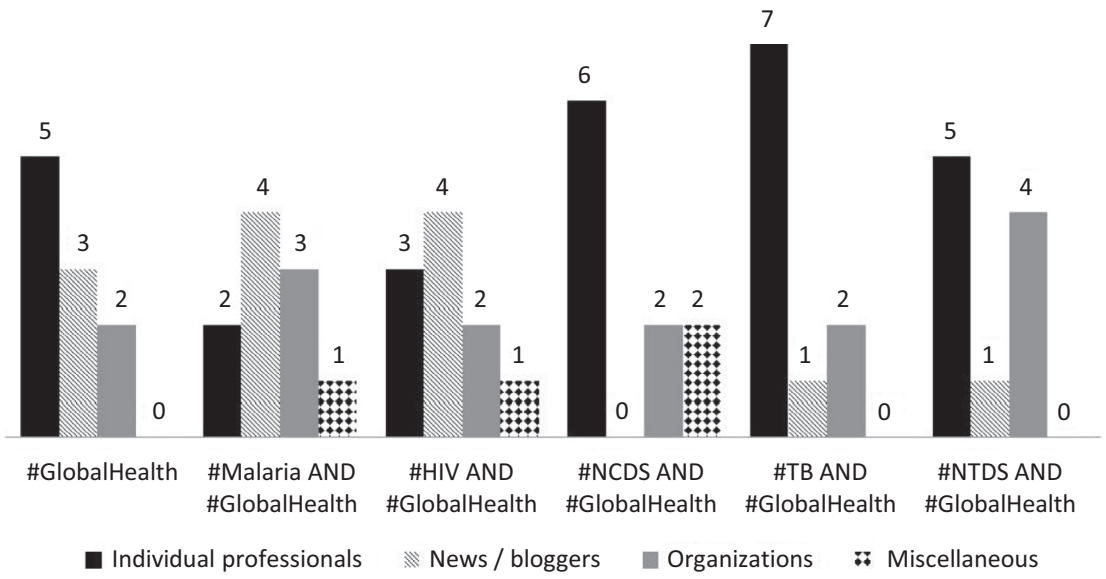

Figure 1. **Top 10 Twitter users in the #GlobalHealth corpus and in the 5 subcorpora. The users were categorized as individual professionals (black), news sources or bloggers (light downward diagonal), nongovernmental/governmental organizations (gray), and miscellaneous (solid diamond). The numbers on top of the bar are the counts of users. NCDS, noncommunicable diseases; NTDS, neglected tropical diseases; TB, tuberculosis.**

and miscellaneous) (Table 3). Across 5 subcorpora, the most common themes were prevention, control, and treatment (4339/13,087, 33%), followed by advocacy (3706/13,087, 28%). Epidemiological information (1803/13,087, 14%) and societal impact (1617/13,087, 12%) made up a quarter of the tweets. The rest did not fall into any categories that were defined a priori (miscellaneous, 1622/13,087, 12%).

We found differences in the proportion of these 5 categories across the 5 subcorpora (Fisher's exact test, $P < .001$). A third of the #Malaria subcorpus was about advocacy (1148/3640, 32%), and a third was on prevention, control, and treatment (1223/3640, 34%). Nearly half of the #HIV subcorpus was about prevention, control, and treatment (1699/3557, 48%). One-half of the #NCDS subcorpus was for advocacy for noncommunicable diseases (1186/2373, 50%). A third of the #TB subcorpus was on prevention, control, and treatment (562/1781, 32%), and a quarter was on societal impact (453/1781, 25%). In the #NTDS subcorpus, about 6 in 10 tweets (990/1736, 58%) were about advocacy for neglected tropical

Table 3. **Distribution in Themes of Topics of Tweets of 5 Subcorpora With Hashtag #GlobalHealth and 1 of the Following Hashtags: #Malaria, #HIV, #NCDS, #TB, and #NTDS**

	#Malaria and #GlobalHealth	#HIV and #GlobalHealth*	#NCDS and #GlobalHealth	#TB and #GlobalHealth	#NTDS and #GlobalHealth	TOTAL
	n (%)	n (%)	n (%)	n (%)	n (%)	n (%)
Advocacy	1148 (31.54)	0 (0)	1186 (49.98)	382 (21.45)	990 (57.03)	3706 (28.32)
Epidemiological Information	390 (10.71)	695 (19.54)	175 (7.37)	308 (17.29)	235 (13.54)	1803 (13.78)
Prevention, Control and Treatment	1223 (33.60)	1699 (47.76)	592 (24.95)	562 (31.56)	263 (15.15)	4339 (33.16)
Societal Impact	0 (0)	744 (20.92)	420 (17.70)	453 (25.44)	0 (0)	1617 (12.36)
Miscellaneous	879 (24.15)	419 (11.78)	0 (0)	76 (4.27)	248 (14.29)	1622 (12.39)
TOTAL	3640 (100)	3557 (100)	2373 (100)	1781 (100)	1736 (100)	13087

NCDS, noncommunicable diseases; NTDS, neglected tropical diseases; TB, tuberculosis.

* After removing sparse terms that only appeared 3 or fewer times in the corpus from the document-term matrix, 1 tweet was left with no terms in the document-term matrix and was therefore excluded from the Latent Dirichlet Allocation model. That tweet was written in German but with English hashtags #globalhealth and #HIV.

Annals of Global Health, VOL. 83, NO. 3–4, 2017
May–August 2017: 682–690

Table 4. Example Tweets Under One of the Content Themes in the Subcorpus of #TB and #GlobalHealth

	Example tweet			
Theme	Twitter User	Twitter Handle	Body of the Tweet	Link to the Tweet
Advocacy	Skoll Foundation	@SkollFoundation	Great #WorldTBDay story with vivid photographs about how @HeroRATs sniff #TB http://dailym.ai/1N5E6D8 #socent #globalhealth @AFP	https://twitter.com/SkollFoundation/status/580478789156806656
Epidemiological information	USAID Global Health	@USAIDGH	#TB is a major #globalhealth concern, killing 1.3M ppl/yr & infecting 8.6M, despite being preventable&curable	https://twitter.com/USAIDGH/status/431870139639009280
Prevention, control, and treatment	Eli Lilly & Company	@LillyPad	Study finds 86% lower patient mortality if #TB & #HIV are treated together instead of separately. #globalhealth	https://twitter.com/LillyPad/status/435822388245438464
Societal impact	USAID Global Health	@USAIDGH	Globally, #TB is strongly associated with #poverty & poor living conditions http://ow.ly/ueTye #endpoverty #globalhealth	https://twitter.com/USAIDGH/status/441628196589826048
Miscellaneous	Pulitzer Center	@pulitzercenter	Our grantee @Meera_Senthi reports from South Africa abt drug-resistant #TB epidemic. Watch this video: http://bit.ly/TBinsouthafrica #globalhealth	https://twitter.com/pulitzercenter/status/521793884540907520

TB, tuberculosis; USAID, United States Agency for International Development.

diseases. In Table 4, a tweet that is representative of each theme from the #TB subcorpus is provided as an example. For example, Twitter users might discuss treatment, a tweet that falls into "prevention, control, and treatment" theme, as follows: "Study finds 86% lower patient mortality if #TB & #HIV are treated together instead of separately. #globalhealth." Likewise, Twitter users might campaign for a specific cause—in this case, promoting the World TB Day, "Great #WorldTBDay story with vivid photographs about how @HeroRATs sniff #TB http://dailym.ai/1N5E6D8 #socent #globalhealth @AFP." For detailed topic modeling results of the Latent Dirichlet Allocation analysis, see the Online Supplementary Materials.

In the Online Supplementary Materials, the time trend of the daily frequency of tweets in the #GlobalHealth corpus (Supplementary Fig. 1; see the Online Supplementary Materials at doi:10.1016/j.aogh.2017.09.006), and the 5 subcorpora (Supplementary Figs. 2, 3, 4, 5, and 6; see the Online Supplementary Materials at doi:10.1016/j.aogh.2017.09.006) are presented. The news and events that triggered the spikes of Twitter activity are also presented in the Online Supplementary Materials.

DISCUSSION

This study provides a snapshot of global health–related Twitter conversations specific to 5 major global health issues (HIV, tuberculosis, malaria, neglected tropical diseases, and noncommunicable diseases) from January 2014 to April 2015. Our results highlighted that, overall, information about prevention, control, and treatment as well as advocacy predominated the #GlobalHealth Twitter conversations that were specifically pertinent to the diseases under study here.

Campaigns against the "big 3" infectious diseases—HIV/AIDS, tuberculosis, and malaria—that had been supported by the Global Fund have long been dominating the global health conversation.[13,14] Since the early 21st century, advocates for research and intervention against a group of lesser known infectious diseases, collectively known as neglected tropical diseases, have made these diseases high in the agenda of global health.[15] Noncommunicable diseases have also been advocated in more recent global health discussion, as emerging economies, such as China and India, have been experiencing the epidemiological transition

wherein prevalence of cancer and other noncommunicable diseases are on the rise.[16] The higher percentage of tweets in advocacy in the #NCDS and #NTDS subcorpora may reflect the reality that within the discourse of global health, these 2 groups of diseases still need advocates. The "big 3" are so well established in the discourse of global health that the need to advocate for them within the global health community is less prominent.

We described the top 10 users for the #GlobalHealth corpus, and those for the subcorpora of #Malaria, #HIV, #NCDS, #TB, and #NTDS. Although none of the 3 categories dominated the lists of top 10 users, it is interesting to note that some individual professionals are highly active in the global health communication on Twitter. Some of them are journalists specializing in global health issues; others are public health professionals and academics who turned to Twitter to advocate for global health. Future research should investigate the role of these potential Twitter key opinion leaders on shaping public opinion on matters pertinent to global health.

The identification of news and events that triggered elevated Twitter activity in the #GlobalHealth corpus and the subcorpora gave us hints into what types of issues might attract attention of Twitter users who are interested in global health.

Our study has several limitations. First, our analysis was limited to a corpus of tweets with the hashtag #GlobalHealth, and thus we excluded tweets with contents on global health issues without the use of the hashtag #GlobalHealth. However, by limiting our analysis to #GlobalHealth tweets, our study was more specific because only Twitter users who wanted to emphasize the concept of global health would likely use the hashtag #GlobalHealth. This would reduce the noise introduced by irrelevant tweets. Likewise, our decision to focus on 5 subcorpora with both #GlobalHealth and another disease-specific hashtag reduced our sample to small subsets of tweets that specifically discussed such diseases in the context of global health. Our purpose was to limit our analysis to those tweets with an explicit emphasis of the diseases (through using specific hashtags) in the context of "global health" instead of general discussion of such diseases.

Second, given our choice of hashtags that were English based, our corpus of tweets was predominately in English (Table 1). Thus, our results describe the Twitter conversation on global health by users from the English-speaking world as well as the English-speaking educated class in the rest of the world. We did not explicitly exclude non-English tweets from our analysis of the corpus, because given the low prevalence of other languages in the corpus, the positive predictive value of the Twitter metadata of the language of the tweet being something other than English was expected to be low. For example, in our corpus of #GlobalHealth tweets, 188 tweets were labeled as Danish ("da") by Twitter, of which 168 were actually in English (manually coded by the corresponding author). Taking into account the possibility of misclassification of English tweets as non-English tweets, we chose to include all the tweets within the corpus and the subcorpora in our topic models. We acknowledge the presence of a small percentage of genuinely non-English tweets therein.

Third, misclassification of tweets into alternate themes was a possibility, because we used Latent Dirichlet Allocation models to first categorize tweets into topics before we manually categorized the topics into themes. However, this method was significantly faster than manual coding of thousands of tweets. Because our goal was to obtain an overall picture, we struck a reasonable balance between efficiency and accuracy by adopting this method.

Fourth, we used the time zones as an indicator to the approximate location of the users, because the majority of tweets do not carry geolocation data. Although this can be seen as a limitation, the time zone data provide us with a big picture of where these tweets originated. It is no surprise that in a corpus of tweets that are predominantly written in English, the majority of the users used time zones of North America or the United Kingdom.

Fifth, given the cross-sectional study design, changes over time cannot be evaluated, and tweets after the end of April 2015 were beyond the scope of this paper.

In conclusion, this study described a corpus of tweets with the hashtag #GlobalHealth, in which the contents of 5 subcorpora with co-occurring hashtags, #Malaria, #HIV, #TB, #NCDS, and #NTDS, were analyzed with greater details. Across the 5 subcorpora, the most common theme was prevention, control, and treatment, followed by advocacy. Among the top 10 users who tweeted the highest number of tweets with #GlobalHealth, and in the 5 subcorpora with co-occurring hashtags, were primarily individual professionals (journalists and scientists), governmental or nongovernmental organizations, and news media. The most common time zones of users in our #GlobalHealth corpus were those in North America or the United Kingdom. Our descriptive study laid the foundation to future in-depth research of global health conversations on Twitter.

Annals of Global Health, VOL. 83, NO. 3–4, 2017
May–August 2017: 682–690

ACKNOWLEDGMENTS

I.C.H.F. thanks Dr. Chung-Hong Chan of the Journalism and Media Studies Centre at The University of Hong Kong for initially teaching him topic modeling and providing him with sample R codes for topic modeling.

SUPPLEMENTARY DATA

Supplementary tables accompanying this article can be found in the online version at doi:10.1016/j.aogh.2017.09.006.

REFERENCES

1. GBD 2015 Mortality and Causes of Death Collaborators. Global, regional, and national life expectancy, all-cause mortality, and cause-specific mortality for 249 causes of death, 1980-2015: a systematic analysis for the Global Burden of Disease Study 2015. Lancet 2016;388:1459–544.
2. Koplan JP, Bond TC, Merson MH, et al. Towards a common definition of global health. Lancet 2009;373:1993–5.
3. Greenwood S, Perrin A, Duggan M. Social Media Update 2016: Facebook Usage and Engagement is on the Rise, while Adoption of other Platforms Holds Steady. Washington, DC: Pew Research Center; 2016.
4. Fung ICH, Duke CH, Finch KC, et al. Ebola virus disease and social media: a systematic review. Am J Infect Control 2016;44:1660–71.
5. Fu KW, Liang H, Saroha N, Tse ZTH, Ip P, Fung IC-H. How people react to Zika virus outbreaks on Twitter? A computational content analysis. Am J Infect Control 2016;44:1700–2.
6. Sharma M, Yadav K, Yadav N, Ferdinand KC. Zika virus pandemic—analysis of Facebook as a social media health information platform. Am J Infect Control 2016;45:301–2.
7. Twitter. Available at: https://about.twitter.com/company. Accessed March 6, 2017.
8. Prue CE, Lackey C, Swenarski L, Gantt JM. Communication monitoring: shaping CDC's emergency risk communication efforts. J Health Comm 2003;8:35–49.
9. Bedrosian SR, Young CE, Smith LA, et al. Lessons of risk communication and health promotion—West Africa and United States. MMWR Suppl 2016;65(Suppl 3):68–74.
10. Blei DM. Probabilistic topic models. Communicat ACM 2009;55:77–84.
11. Citizens News Service. About us. Lucknow, India: Citizens News Service. Available at: http://www.citizen-news.org/2007/07/about-us.html. Accessed April 13, 2017.
12. Deutsche Stiftung Weltbevolkerung. #LetsSaveLives: If you do one thing today, support global health research and development. Hannover, Germany: DSW; 2014. Available at: http://www.dsw.org/en/2014/10/letssavelives-support-global-health-research-development/. Accessed March 8, 2017.
13. Hanefeld J. The global fund to fight AIDS, tuberculosis and malaria: 10 years on. Clin Med (Lond) 2014;14:54–7.
14. Tan DH, Upshur RE, Ford N. Global plagues and the Global Fund: challenges in the fight against HIV, TB and malaria. BMC Int Health Hum Rights 2003;3:2.
15. Hotez P. A new voice for the poor. PLoS Negl Trop Dis 2017;1:e77.
16. Bollyky TJ, Emanuel EJ, Goosby EP, Satcher D, Shalala DE, Thompson TG. NCDs and an outcome-based approach to global health. Lancet 2014;384:2003–4.