Twitter as a Tool for Health Research: A Systematic Review

Lauren Sinnenberg, BA, Alison M. Buttenheim, PhD, MBA, Kevin Padrez, MD, Christina Mancheno, BA, Lyle Ungar, PhD, and Raina M. Merchant, MD, MSHP

Background. Researchers have used traditional databases to study public health for decades. Less is known about the use of social media data sources, such as Twitter, for this purpose.

Objectives. To systematically review the use of Twitter in health research, define a taxonomy to describe Twitter use, and characterize the current state of Twitter in health research.

Search methods. We performed a literature search in PubMed, Embase, Web of Science, Google Scholar, and CINAHL through September 2015.

Selection criteria. We searched for peer-reviewed original research studies that primarily used Twitter for health research.

Data collection and analysis. Two authors independently screened studies and abstracted data related to the approach to analysis of Twitter data, methodology used to study Twitter, and current state of Twitter research by evaluating time of publication, research topic, discussion of ethical concerns, and study funding source.

Main results. Of 1110 unique health-related articles mentioning Twitter, 137 met eligibility criteria. The primary approaches for using Twitter in health research that constitute a new taxonomy were content analysis (56%; n = 77), surveillance (26%; n = 36), engagement (14%; n = 19), recruitment (7%; n = 9), intervention (7%; n = 9), and network analysis (4%; n = 5). These studies collectively analyzed more than 5 billion tweets primarily by using the Twitter application program interface. Of 38 potential data features describing tweets and Twitter users, 23 were reported in fewer than 4% of the articles. The Twitter-based studies in this review focused on a small subset of data elements including content analysis, geotags, and language. Most studies were published recently (33% in 2015). Public health (23%; n = 31) and infectious disease (20%; n = 28) were the research fields most commonly represented in the included studies. Approximately one third of the studies mentioned ethical board approval in their articles. Primary funding sources included federal (63%), university (13%), and foundation (6%).

Conclusions. We identified a new taxonomy to describe Twitter use in health research with 6 categories. Many data elements discernible from a user's Twitter profile, especially demographics, have been underreported in the literature and can provide new opportunities to characterize the users whose data are analyzed in these studies. Twitter-based health research is a growing field funded by a diversity of organizations.

Public health implications. Future work should develop standardized reporting guidelines for health researchers who use Twitter and policies that address privacy and ethical concerns in social media research. (The full article is available online. *Am J Public Health.* 2017;107: 143, e1–e8. doi:10.2105/AJPH.2016.303512)

PLAIN-LANGUAGE SUMMARY

Twitter is an interactive social media platform established in 2006 that allows users to send 140-character messages to one another. Public health researchers have begun to use Twitter for research—both to interact with study participants and to mine the platform for data. This growing body of work, however, has not yet been systematically studied. In this review, we analyzed 137 studies that used Twitter to conduct health research that collectively analyzed more than 5 billion tweets. We found that the majority of articles (57%) focused on analyzing the content of tweets, whereas other studies harnessed Twitter's interactive features for recruitment or interventions. Most studies were published in the past 2 years, and were supported by a wide variety of funders. Twitter-based public health research is a growing field. Future work is needed to help create standardized reporting guidelines to improve the reproducibility and comparability of Twitter studies.

One of the 3 main functions of public health is the "assessment and monitoring of the health of communities and populations at risk to identify health problems and priorities."[1(p48)] For decades, health researchers have leveraged large databases of health information for this purpose. In recent years, researchers have recognized that social media platforms, such as Twitter, Facebook, and Instagram, can also provide data about population-level health and behavior.[2–6]

Among social media networks, Twitter provides a unique big data source for public health researchers because of the real-time nature of the content, and the ease in accessing and searching publically available information. The reach and volume of data are also significant—every day, 500 million tweets are sent by more than 300 million active users worldwide.[7] Although Twitter users are not representative of the population of the United States (persons aged < 50 years and dwelling in urban areas are most likely to use Twitter[8]), a wide breadth of demographic groups is represented. In addition to its potential as a more traditional data source, Twitter is also interactive; researchers can contribute to the social network and harness this feature as a recruitment tool or for an intervention.

Despite the potential for this social media platform, the landscape of how Twitter is and might be used for health research has yet to be defined.[9–24] There is value in understanding the ways that the Twitter data set can be harnessed to contribute to our understanding of public health.

We sought to systematically review the literature of health-related research that used Twitter. We focused primarily on characterizing these research studies and developing a taxonomy. We also evaluated other features of Twitter health research including how researchers accessed Twitter information, which Twitter data features researchers reported, and measures of the current state of health research using Twitter. This review can provide insights about Twitter-based research and identify new opportunities for assessing and monitoring health on social media platforms.

METHODS

This study was a systematic review of Twitter-based health research according to the PRISMA statement.[25]

We performed a systematic literature search on July 8, 2015, by searching PubMed, Embase, Web of Science, and CINAHL for articles whose title, abstract, or search keywords matched the following Boolean search strings: "Twitter AND health," "tweet AND health," and "Twitter AND medicine." We also searched Google Scholar by using the previously stated search strings and reviewed it for unique articles. We performed a second search on September 8, 2015, by searching the previously mentioned databases for the following additional Boolean search strings: "Twitter AND illness" and "Twitter AND disease."

We included studies if they met the following criteria: (1) primary peer-reviewed journal article representing original health research, (2) methodology and results provided, and (3) Twitter was used by researchers to obtain at least part of the results. We defined health research as research that contributes to the World Health Organization's definition of health: "a state of complete physical, mental and social well-being and not merely the absence of disease or infirmity."[26] We excluded abstracts, editorials, review articles, and non-English articles.

We screened articles that met inclusion criteria for quality by using a modified Quality Assessment Tool for Systematic Reviews of Observational Studies.[27] We assessed studies on the basis of 4 quality criteria: presence of a clearly defined objective, description of techniques for extraction of Twitter data (if applicable), discussion of coding methodology including intercoder reliability, and discussion of limitations. We excluded articles with a low rating on 2 or more of the quality criteria from the review. Two authors independently screened studies for inclusion. Disagreements were discussed and adjudicated by consensus. The κ score between 2 coders for a 10% sample for study inclusion was 0.74.

To develop and define the taxonomy of Twitter use, we developed a codebook based on the studies that met inclusion criteria that described the manner in which health researchers have used Twitter in their work. Four authors (L. S., A. M. B., C. M., R. M. M) developed a preliminary codebook by including 14 themes that were then combined and revised to compose the 6 themes reported in this review. We further categorized the 6 themes into 2 subsets: studies that analyzed data from Twitter and studies that used the Twitter social platform. Two authors coded all inclusion articles, adjudicating differences with a larger group that included 2 other authors. We counted studies that used Twitter in more than 1 way in all applicable themes.

To characterize approaches for accessing Twitter data, we reviewed each article for its method of mining Twitter. Every tweet can potentially generate 38 data features related to the Twitter user (e.g., age, occupation, socioeconomic status) and the tweet (e.g., timing, location, content, sentiment, language; Table 1). Two authors (L. S., C. M.) identified these 38 data features in an iterative process based on the data features reported in the studies reviewed in this article, as well as data features identified in a literature search. We assessed each article for its reporting of these specific data features.

To describe the current state of Twitter health research, we extracted the following information from studies in this review: research field, research topic, publication year, and funding source. We also extracted information about the institutional review board or ethical review board process associated with accessing and using Twitter-based data. We collected and coded the following additional elements but did not report them in this article: journal, type of

ABOUT THE AUTHORS

Lauren Sinnenberg, Kevin Padrez, Christina Mancheno, and Raina M. Merchant are with Penn Medicine Social Media and Health Innovation Lab, University of Pennsylvania, Philadelphia. Lyle Ungar is with the Department of Computer and Information Science, University of Pennsylvania. Alison M. Buttenheim is with Department of Family and Community Health, School of Nursing, University of Pennsylvania, and Center for Health Incentives and Behavioral Economics, University of Pennsylvania.

Correspondence should be sent to Raina M. Merchant, MD, MSHP, Department of Emergency Medicine, Penn Medicine Social Media and Health Innovation Lab, 423 Guardian Dr, University of Pennsylvania, Philadelphia, PA 19104 (e-mail: raina.merchant@uphs.upenn.edu). Reprints can be ordered at http://www.ajph.org by clicking the "Reprints" link.

This article was accepted October 8, 2016.

doi: 10.2105/AJPH.2016.303512

TABLE 1—Frequency of Use of Different Metadata Elements That Can Be Extracted From Twitter in Included Articles: 2010–2015

Data	Articles, No. (%)
Explicit data about user	
Twitter handle	6 (4)
Language	0 (0)
Time zone	4 (3)
Location	12 (9)
Date account created	5 (4)
User profile (free text)	5 (4)
User profile photo	2 (1)
Total number of tweets	11 (8)
Number of followers	26 (19)
Number following	6 (4)
Extractable data about user	
Age	1 (1)
Gender	5 (4)
Marital status	1 (1)
Political party	0 (0)
Race/ethnicity	1 (1)
Occupation	5 (4)
Interests	0 (0)
Religion	0 (0)
Income	1 (1)
Mood	1 (1)
Disease state	5 (4)
Network	5 (4)
Explicit data from tweet	
140 characters	112 (82)
#hashtag	13 (9)
URL	6 (4)
Geotag	28 (20)
Application used to send tweet	2 (1)
Number of retweets	12 (9)
Number of favorites	1 (1)
Linked images	2 (1)
User mentions	3 (2)
Time and date of tweet	19 (14)
Extractable data about tweet	
Content	77 (55)
Sentiment	21 (15)
Image analysis	1 (1)
Language	4 (3)

Note. URL = uniform resource locator.

journal, journal impact factor, objectives, nation in which study was conducted, population studied, and timeframe of tweets studied.

We used summary statistics to quantify the frequency of themes used across studies, methodologies of accessing Twitter, Twitter data features reported, and measures of the current state of Twitter health research. We compiled and analyzed summary data in Stata SE version 14 (StataCorp LP, College Station, TX).

RESULTS

Of the 3049 articles identified in the initial database search, we found 1110 articles to be unique. Of these, 137 of these articles met eligibility criteria (Figure A, Figure B, and Table E, available as supplements to the online version of this article at http://www.ajph.org).

We identified 4 ways in which Twitter data were used by health researchers in their work: content analysis of tweets (56%; n = 77), surveillance of volume of tweets about a particular topic (26%; n = 36), engagement of Twitter users with Twitter accounts or tweets (14%; n = 19), and network analysis of Twitter users (4%; n = 137) and 2 ways that the Twitter platform was used for research: recruitment of participants for research (7%; n = 9) and Twitter-based intervention (7%; n = 9; Table 2). We included studies that used more than 1 methodology for studying Twitter (12%; n = 17) in all applicable themes.

Twitter as a Data Source

Content analysis. Of the articles included in this review, more than half (56%; n = 77) analyzed the content of tweets about a specific health topic to characterize public discourse on Twitter. Within this group, there were subcategories including sentiment analysis (15%; n = 21) and image analysis (1%; n = 1). This article analyzed the content of images attached to tweets with the hashtag "thinspo" to assess the types of images that are used by individuals to discuss body image.[28]

Surveillance. Many of the articles (26%; n = 36) analyzed Twitter data for frequency of discussion of a particular topic. The majority of the articles in this group were focused on monitoring Twitter for mentions of influenza-related terms compared with the normal background discussion of influenza. Of the 36 articles in this category, 19% (n = 7) utilized these changing frequencies to predict outcomes such as asthma emergency department visits and heart disease mortality.

Engagement. Several articles (14%; n = 19) assessed Twitter presence as well as user interactions with content produced by other users. Researchers with articles in this category reported metrics including number of retweets and number of favorites to measure how the Twitter community responded to tweets from different users or to tweets about different health topics. Six of these studies examined Twitter usage by health organizations and how the public engages with tweets from those organizations.[29–34]

Network analysis. The smallest category (4%; n = 5) of articles analyzed the networks of Twitter users. Two of these studies specifically assessed the connections between patients with cancer on Twitter and sought to identify the hubs, or Twitter users with the greatest connectedness to other users within the network.[35,36]

Utilization of the Twitter Platform

Recruitment. Twitter was used to recruit research survey participants in 7% (n = 9) of the studies. In 1 article, researchers were able to reach a highly unique population (owners of venomous snakes) by recruiting via Twitter.[37]

Intervention. Some of the studies (7%; n = 9) used Twitter as an intervention in their study. Two of these studies investigated the effectiveness of weight loss interventions with a Twitter-based social component.[38,39] Many of the studies in this group studied behavioral aspects of health issues such as weight loss, smoking, or nutrition.

Use of Twitter Data

Accessing Twitter data. Of the 137 articles in this review, 108 analyzed tweets (those that did not were largely studies that primarily utilized the Twitter platform for their research). These 108 articles represented more than 5.1 billion analyzed tweets (Table A, available as a supplement to the online version of this article at http://www.ajph.org). Many studies (41%; n = 44) used the Twitter application programing interface (API) to mine Twitter, and others included NodeXL, Topsy, and NCapture (Table B, available as a supplement to

TABLE 2—Taxonomy of Use of Twitter-Generated Data in Included Articles: 2010–2015

Taxonomy	Description	Articles, No. (%)	Examples
Use of Twitter data			
Content analysis	Assessment of body of tweets for themes in relation to a specific subject	77 (56)	Smoking, diabetes, obesity, concussion
Sentiment analysis	Assessment of body of tweets for positive or negative discussion of a specific subject	21 (15)	Schizophrenia, vaccination, trans health
Image analysis	Assessment of images within body of tweets for themes in relation to a specific subject	1 (1)	#thinspo
Surveillance	Monitoring of Twitter traffic for mentions of a particular topic above the normal background level of discussion	36 (26)	Influenza, Ebola, adverse drug reactions
Prediction	Using Twitter to estimate prevalence of disease or behavior	7 (5)	Heart disease mortality, influenza infection, Affordable Care Act enrollment, asthma emergency department visits
Engagement	Assessing impact of discussion on Twitter by analyzing presence of an account, number of retweets, favorites, followers, etc.	19 (14)	Nutrition public health marketing campaign, social media impact of local health departments, social media adoption by pharmaceutical companies
Network analysis	Assessing the relationship and interactions between Twitter users about a certain topic	5 (4)	Communities of cancer patients, sharing of health information by health organizations
Use of Twitter platform			
Recruitment	Use of Twitter to enroll patients in research studies	9 (7)	Recruit for surveys of youth soccer parents, keepers of venomous snakes, smokers
Intervention	Use of Twitter as an intervention in a research study	9 (7)	Weight loss randomized controlled trial, smoking cessation

the online version of this article at http://www.ajph.org).

Twitter metadata. Of the 38 potential Twitter data features extracted for review, 25 data features were reported by less than 5% of studies. The data most often included in analysis were content analysis (55%; n = 77), geotags (20%; n = 28), and number of followers (19%; n = 26). Very few studies (4%; n = 5), reported on the demographics of the Twitter users producing the content used in their research including age, gender, and race (Table 1).

Current State of Twitter Health Research

Publication date. Most articles were recent, with 33% (n = 46) published in 2015 compared with only 1% (n = 2) published in 2010, the first year represented in this review (Figure 1).

Research field and topic. The most commonly represented research fields in the review were public health (22%; n = 30), infectious disease (20%; n = 27), behavioral medicine (18%; n = 24), and psychiatry (11%; n = 16; Table 3 and Table F, available as a supplement to the online version of this article at http://www.ajph.org). The most common research topics included influenza (8%; n = 11), smoking (7%; n = 9), cancer (5%; n = 7), and Ebola (4%; n = 5; Table 3).

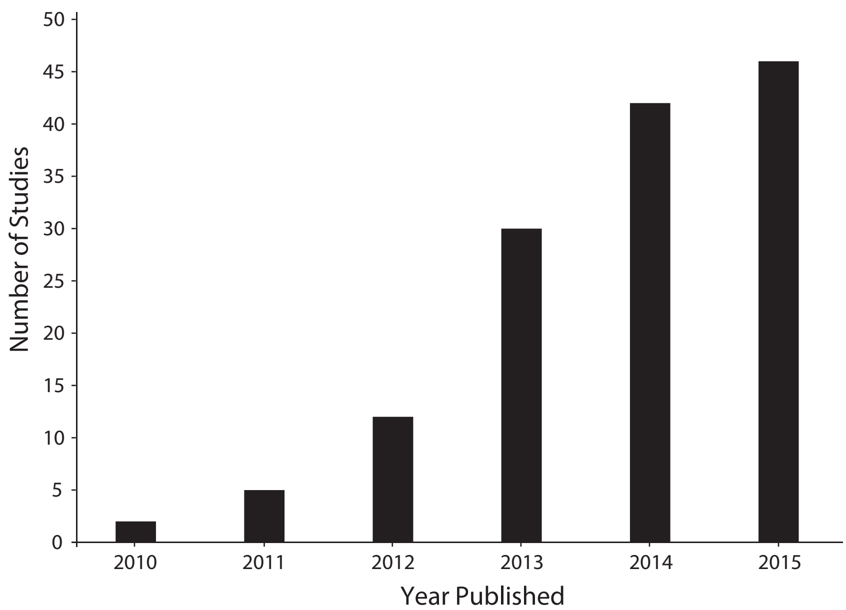
Ethical discussion. Several articles (32%; n = 43) discussed acquiring ethics board approval for their research. Fewer articles (12%; n = 16) discussed consent for use of the Twitter data. The majority of articles that discussed consent (56%; n = 9 of 16) used traditional online or written consent with study participants. These 9 studies all involved use of Twitter as an intervention or for participant recruitment. Several articles (4%; n = 6) accounted for consent via the Twitter terms of service or by stating that all data published on Twitter is public (Table C, available as a supplement to the online version of this article at http://www.ajph.org).

Funding. There was a wide breadth of funding organizations supporting the studies included in this review. Funding by the National Institutes of Health represented 38% of the funding sources reported by the articles. Many other federal organizations, foundations, and universities were also identified as funding Twitter-based research (Table D, available as a supplement to the online version of this article at http://www.ajph.org).

DISCUSSION

This review has 3 main findings. First, we defined a new taxonomy to describe how Twitter is used in health research consisting of 6 categories: content analysis, surveillance, engagement, recruitment, intervention, and network analysis. Second, we identified a significant amount of variability in how Twitter use was reported in the studies, and many types of Twitter-based data that have to date been underutilized by health researchers. Finally, we described the current state of Twitter in health research and found a growing field as evidenced by the increasing number of Twitter-related publications each year and the diversity of funding organizations.

We defined a new taxonomy describing 6 different uses of Twitter among health researchers. The studies in our review were varied and included using Twitter to analyze stigma associated with schizophrenia,



Note. This figure shows the number of articles in this systematic review published each year from 2010 to 2015. The x-axis shows each year that an inclusion article was published. The y-axis demonstrates the number of studies in each publication year.

FIGURE 1—Publication Date of Included Articles

to predict enrollment in the Health Insurance Exchanges based on sentiment about the Affordable Care Act, to recruit difficult-to-reach populations for public health–related surveys, and to create communities to help patients with weight loss.[37–41] This taxonomy may help to delineate the opportunities for health researchers to utilize Twitter data or the Twitter platform in their work. Each of these approaches allows health researchers to access groups of individuals and large quantities of data about the public with relative efficiency and presumably low cost compared with traditional public health databases.

Underuse of Twitter Metadata

The most common approach to Twitter mining was via the Twitter API. This is a free application that allows access to 1% of all Twitter data in real time. The use of the Twitter API highlights one of the strengths of the Twitter database; it allows free access to large public data immediately after the data are created. Fewer studies leveraged the Twitter Firehose, which provides access to the entire Twitter data set. Gaining access to the Firehose can be costly; however, it may be necessary to improve quantity and quality of data for researchers investigating a rare disease topic.

The articles in our study represent the analysis of more than 5.1 billion tweets. Each tweet can potentially generate 38 data features including detailed metadata about the Twitter user and tweet. Although some data including the content of the tweet and the geotag of the tweet were reported by several studies, many data featured were not included in data analysis. In particular, very few articles addressed the demographics of the Twitter users whose tweets they analyzed. In the context of public health, it is critically important to understand the different populations and communities contributing to the discourse on social media. This suggests a missed opportunity for better characterizing Twitter users and the ways in which their demographic attributes affect their participation in the network, and how both of these affect their health status.

Analysis of Twitter demographics can be difficult because it cannot be directly obtained from the data source as Twitter does not collect this information nor report it. Previous reports have shown, however, that demographic information including age, gender, socioeconomic status, religion, and personality type can be extrapolated from a user's tweets via machine learning, with accuracy ranging from 60% to 90%.[42–46] This is an underutilized resource in Twitter-based research that can be applied for richer information about the Twitter users via interdisciplinary work with computer and data scientists.

Previous studies have analyzed Twitter and other social media data sources relative to the cognate constructs and domains of these data types.[9–12,14–24] Many of these studies have discussed the constraints of such media and potential solutions to manage the limitations inherent in social media data analyses. These include but are not limited to using computational linguistic methods to evaluate signal from noise (e.g., sentiment analysis, lexical metrics, linguistic inquiry and word count analyses), applying geospatial analytics to extract location information, analyzing link-based (i.e., relational, social network) rather than point-based (i.e., individual) data, selection of streaming versus search API—versus accessing the "Firehose," comparative issues associated with using tweets versus retweets, or tweets with or without URLs, and methods to assess real space (i.e., gold standard) validity.[9,47–52] Although our review did not specifically code for all of these variables, these previously described approaches speak to the broader potential and opportunities for using social media data for public health research and interventions.

In the 137 articles included in this study, there was great diversity in their methodological discussion of their Twitter use. There is a clear need for Twitter research reporting standards that will allow better comparisons between Twitter-based studies, improve the ability for replication, and add clarity to the understanding of the methodology of this research.

Future Directions in Twitter Research

Twitter-based health research is a rapidly growing field. In our study, each year showed approximately a two-fold increase in number of publications. We also identified that Twitter research is supported by

TABLE 3—Most Common Research Fields for Included Articles: 2010–2015

Research Field	Articles, No. (%)	Research Topics
Public health	30 (22)	Affordable Care Act, health organizations, obesity, pet exposure, sexual health, transgender health, vaccination
Infectious disease	27 (20)	Antibiotics, cholera, Ebola, enterohemorrhagic *Escherichia coli*, HIV, influenza, measles, sexually transmitted infections
Behavioral medicine	24 (18)	Nutrition, physical activity, sleep disorders, smoking, weight loss
Psychiatry	15 (11)	Anorexia, bipolar disorder, depression, drug abuse, emotions, marijuana, mental health, obsessive–compulsive disease, schizophrenia, stimulant use, suicide
Neurology	9 (7)	Concussion, deafness, dementia, epilepsy, migraine, multiple sclerosis
Oncology	6 (4)	Cancer
Obstetrics and gynecology	5 (4)	Prenatal health, breastfeeding, cancer, polycystic ovarian syndrome
Dentistry	4 (3)	Dental pain, orthodontics
Pharmacy	4 (3)	Adverse drug reactions, online pharmaceutical presence
Emergency medicine	3 (2)	Asthma, cardiac arrest, emergency medical services
Pediatrics	2 (1)	Pediatric obesity, health literacy
Endocrinology	2 (1)	Diabetes
Allergy and immunology	1 (1)	Allergy
Anesthesia	1 (1)	Pain
Cardiology	1 (1)	Heart disease mortality
Hematology	1 (1)	Stem cell
Radiology	1 (1)	Radiation

a wide variety of funding organizations including the federal government through the National Institutes of Health, National Science Foundation, the Food and Drug Administration, foundations, and universities. The diversity and breadth of funding opportunities suggest value and broad interest in Twitter-based health research.

The most commonly studied topics included important high morbidity and mortality conditions including influenza, cancer, and Ebola, and health behaviors such as smoking. It is interesting that many of the most prevalent and costly chronic diseases in the United States including diabetes and hypertension were less frequently investigated by the studies included in this review. The Centers for Disease Control and Prevention estimates that 9% and 33% of the US population has diabetes and hypertension, respectively, and together they cost the United States $291 billion in 2012.[53–55] In identifying this contrast between the massive burden of these diseases and their relative scarcity in digitally based research, the National Heart, Lung, and Blood Institute has called for "studies to establish the validity, reliability, and scalability of electronic tools for primary data collection."[56(p365)] Whereas some chronic diseases such as hypertension have been shown to be "under-tweeted" relative to its prevalence, other chronic diseases including diabetes and heart failure have been shown to be "over-tweeted" relative to their prevalence.[57] This may suggest potential uses of Twitter to study chronic conditions including hypertension and diabetes.

As the use of social media as a data source for public health researchers is emerging, policies regarding privacy and consent of the users producing the messages have yet to be universally defined. Notably, several of the articles included in this review discussed acquiring approval from their institution's ethics board for their research. Fewer articles discussed obtaining consent for use of the Twitter data. Some studies cited an exemption from ethical approval and consent because of the public nature of the database, and the language included in the Twitter terms of service agreed upon by all Twitter users.

Previous work has demonstrated the challenges in maintaining privacy and anonymity in social media–based research.[58,59] One 2008 study published "anonymized" Facebook profiles of participants that were later found to be identifiable via cross-referencing of variables presented in the study.[60] Twitter-based research faces similar challenges as the traditional frameworks in place to ensure ethical health research may not apply to this freely, publicly accessible data source that does not explicitly contain protected health information. There is clearly a need for universal guidelines addressing ethical social media research, with a focus on protecting the privacy of social media users.

Limitations

This study has several limitations. The search engines we used are primarily meant to catalog peer-reviewed medical journal articles. The use of other search engines or sources such as PsycINFO, Communication and Mass Media Complete, and computer science–focused conference proceedings may have yielded additional results from fields that may have been underrepresented in the search engines we utilized, including communications and computer science. Our search terms were broad in that they used terms such as "medicine," "health," "disease," and "illness" along with "Twitter" to identify Twitter-based health research. Using more disease-specific search terms such as "flu" and "cancer" may have returned additional results that were not included in this review. It is also possible that because of search parameter bias we missed studies that have used Twitter in their research, but used generic terms such as "social media" or "microblogs" rather than the words "Twitter" or "tweet" as a keyword in their title or abstract. We sought to be comprehensive in the data features we reported; however, there may be other items that were not considered in this study.

Future work can describe the specific computational methods used to address the constraints of using social media research including linguistics methods and geospatial analysis, as well as additional data features including type of research (applied [e.g., disease surveillance, quality improvement] and basic [e.g., testing theories of emotional contagion]). We also included only English-language studies. Future work can focus on characterizing research done in other languages.

Conclusions

Twitter is a valuable resource for health researchers interested in capturing live data about a health topic or harnessing the interactive platform for study recruitment or intervention. Twitter-based health research is a growing field as evidenced by the increasing number of publications per year and diversity of funding organizations. This review defined a new taxonomy to describe Twitter use in health research with 6 categories. Many data features that are distillable from a user's Twitter profile, especially demographics, have been underreported in the literature and can provide new opportunities to characterize the users whose data are analyzed in these studies. Future work should develop standardized reporting guidelines for health researchers who use Twitter and policies that address privacy and ethical concerns in social media research. AJPH

CONTRIBUTORS

L. Sinnenberg originated the study, collected and analyzed data, and drafted the article. K. Padrez and C. Mancheno aided in the collection of the data and drafting of the article. A. M. Buttenheim, L. Ungar, and R. M. Merchant aided in the conceptualization of the study and the drafting of the article.

ACKNOWLEDGMENTS

L. Ungar is funded by the Templeton Religious Trust. R. M. Merchant receives funding from National Institutes of Health grants K23 109083 and R01 122457.

HUMAN PARTICIPANT PROTECTION

No institutional review board approval was required for this study as there were no human participants.

REFERENCES

1. World Health Organization Centre for Health Development. A glossary of terms for community health care and services for older persons. 2004. Available at: http://apps.who.int/iris/bitstream/10665/68896/1/WHO_WKC_Tech.Ser._04.2.pdf. Accessed November 4, 2015.

2. King D, Ramirez-Cano D, Greaves F, Vlaev I, Beales S, Darzi A. Twitter and the health reforms in the English National Health Service. *Health Policy*. 2013;110(2-3):291–297.

3. Chew C, Eysenbach G. Pandemics in the age of Twitter: content analysis of tweets during the 2009 H1N1 outbreak. *PLoS One*. 2010;5(11):e14118.

4. Chunara R, Andrews JR, Brownstein JS. Social and news media enable estimation of epidemiological patterns early in the 2010 Haitian cholera outbreak. *Am J Trop Med Hyg*. 2012;86(1):39–45.

5. Neiger BL, Thackeray R, Burton SH, Giraud-Carrier CG, Fagen MC. Evaluating social media's capacity to develop engaged audiences in health promotion settings: use of Twitter metrics as a case study. *Health Promot Pract*. 2013;14(2):157–162.

6. Paul MJ, Dredze M, Michael JP, Mark D. You are what you tweet: analyzing Twitter for public health. *Proceedings of the 5th International AAAI Conference on Weblogs and Social Media*. Menlo Park, CA: The AAAI Press; 2011: 265–272.

7. Company @ About.Twitter.Com. Available at: https://about.twitter.com/company. Accessed November 4, 2015.

8. Duggan M, Brenner J. The demographics of social media users—2012. *Pew Research Center*. 2013:1–14. Available at: http://pewinternet.org/Reports/2013/Social-media-users.aspx. Accessed November 4, 2015.

9. Alshaikh F, Ramzan F, Rawaf S, Majeed A. Social network sites as a mode to collect health data: a systematic review. *J Med Internet Res*. 2014;16(7): e171.

10. Balatsoukas P, Kennedy C, Balatsoukas P, et al. The role of social network technologies in online health promotion: a narrative review of theoretical and empirical factors influencing intervention. *J Med Internet Res*. 2015; 17(6):e141.

11. Charles-Smith L, Reynolds T, Cameron MA, et al. Using social media for actionable disease surveillance and outbreak management: a systematic literature review. *PLoS One*. 2015;10(10):e0139701.

12. Grajales FJ III, Sheps S, Ho K, Novak-Lauscher H, Eysenbach G. Social media: a review and tutorial of applications in medicine and health care. *J Med Internet Res*. 2014;16(2):e13.

13. Hamm M, Chisholm A, Shulhan J, et al. Social media use among patients and caregivers: a scoping review. *BMJ Open*. 2013;3:e002819.

14. Hamm M, Chisholm A, Shulhan J, et al. Social media use by health care professionals and trainees: a scoping review. *Acad Med*. 2013;88(8):1376–1383.

15. Hu Y. Health communication research in the digital age: a systematic review. *J Commun Healthc*. 2015;8(4): 260–288.

16. Kass-Hout T, Alhinnawi H. Social media in public health. *Br Med Bull*. 2013;108(1):5–24.

17. Lardon J, Abdellaoui R, Bellet F, et al. Adverse drug reaction identification and extraction in social media: a scoping review. *J Med Internet Res*. 2015;17(7):e171.

18. Maher C, Lewis L, Ferrar K, Marshall S, De Bourdeaudhuij I, Vandelanotte C. Are health behavior change interventions that use online social networks effective? A systematic review. *J Med Internet Res*. 2014; 16(2):e40.

19. Moorhead S, Hazlett D, Harrison L, Carroll JK, Irwin A, Hoving C. A new dimension of health care: systematic review of the uses, benefits, and limitations of social media for health communication. *J Med Internet Res*. 2013;15(4):e85.

20. Neiger B, Thackeray R, Van Wagenen SA, et al. Use of social media in health promotion purposes, key performance indicators, and evaluation metrics. *Heal Promot Pract*. 2012;13(2):159–164.

21. Sarker A, Ginn R, Nikfarjam A, et al. Utilizing social media data for pharmacovigilance: a review. *J Biomed Inf*. 2015;54:202–212.

22. Sloane R, Osanlou O, Lewis D. Social media and pharmacovigilance: a review of the opportunities and challenges. *Br J Clin Pharmacol*. 2015;80(4):910–920.

23. Velasco E, Agheneza T, Denecke K, Kirchner G, Eckmanns T. Social media and Internet-based data in global systems for public health surveillance: a systematic review. *Milbank Q*. 2014;92(1):7–33.

24. Webb T, Joseph J, Yardley L, Michie S. Using the Internet to promote health behavior change: a systematic review and meta-analysis of the impact of theoretical basis, use of behavior change. *J Med Internet Res*. 2010; 12(1):e4.

25. Moher D. Preferred Reporting Items for Systematic Reviews and Meta-Analyses: the PRISMA statement. *Ann Intern Med*. 2009;151(4):264–269.

26. World Health Organization. WHO definition of "Health." Available at: http://www.who.int/about/definition/en/print.html. Accessed November 4, 2015.

27. Wong WC, Cheung CS, Hart GJ. Development of a quality assessment tool for systematic reviews of observational studies (QATSO) of HIV prevalence in men having sex with men and associated risk behaviours. *Emerg Themes Epidemiol*. 2008;5:23.

28. Ghaznavi J, Taylor LD. Bones, body parts, and sex appeal: an analysis of #thinspiration images on popular social media. *Body Image*. 2015;14:54–61.

29. Veale HJ, Sacks-Davis R, Weaver ER, Pedrana AE, Stoové MA, Hellard ME. The use of social networking platforms for sexual health promotion: identifying key strategies for successful user engagement. *BMC Public Health*. 2015;15:85.

30. Harris JK, Mueller NL, Snider D. Social media adoption in local health departments nationwide. *Am J Public Health*. 2013;103(9):1700–1707.

31. Ramanadhan S, Mendez SR, Rao M, Viswanath K. Social media use by community-based organizations conducting health promotion: a content analysis. *BMC Public Health*. 2013;13(1):1129.

32. Gold J, Pedrana AE, Sacks-Davis R, et al. A systematic examination of the use of online social networking sites for sexual health promotion. *BMC Public Health*. 2011; 11(1):583.

33. Harris JK, Choucair B, Maier RC, Jolani N, Bernhardt JM. Are public health organizations tweeting to the choir? Understanding local health department Twitter followership. *J Med Internet Res*. 2014;16(2):e31.

34. Neiger BL, Thackeray R, Burton SH, Thackeray CR, Reese JH. Use of Twitter among local health departments: an analysis of information sharing, engagement, and action. *J Med Internet Res*. 2013; 15(8):e177.

35. Himelboim I, Han JY. Cancer talk on Twitter: community structure and information sources in breast and prostate cancer social networks. *J Health Commun*. 2014;19(2):210–225.

36. Sugawara Y, Narimatsu H, Hozawa A, Shao L, Otani K, Fukao A. Cancer patients on Twitter: a novel

patient community on social media. *BMC Res Notes*. 2012;5(1):699.

37. van Genderen PJ, Slobbe L, Koene H, Mastenbroek RDL, Overbosch D. Keeping venomous snakes in the Netherlands: a harmless hobby or a public health threat? *Neth J Med*. 2013;71(8):437–441.

38. Turner-McGrievy GM, Tate DF. Are we sure that mobile health is really mobile? An examination of mobile device use during two remotely-delivered weight loss interventions. *Int J Med Inform*. 2014;83(5):313–319.

39. Turner-McGrievy G, Tate D. Tweets, apps, and pods: results of the 6-month Mobile Pounds Off Digitally (Mobile POD) randomized weight-loss intervention among adults. *J Med Internet Res*. 2011;13(4):e120.

40. Joseph AJ, Tandon N, Yang LH, et al. #Schizophrenia: use and misuse on Twitter. *Schizophr Res*. 2015; 165(2-3):111–115.

41. Wong CA, Sap M, Schwartz A, et al. Twitter sentiment predicts Affordable Care Act marketplace enrollment. *J Med Internet Res*. 2015;17(2):e51.

42. Rao D, Yarowsky D, Shreevats A, Gupta M. Classifying latent user attributes in Twitter. In: *Proceedings of the 2nd International Workshop on Search and Mining User-Generated Contents*. New York, NY: Association of Computing Machinery; 2010: 37.

43. Bi B, Kosinski M, Shokouhi M, Graepel T. Inferring the demographics of search users social data meets search queries. In: *Proceedings of the 22nd International Conference on World Wide Web*. New York, NY: Association of Computing Machinery; 2013: 131–140.

44. Pennacchiotti M, Popescu A-M. Democrats, Republicans and Starbucks afficionados: user classification in Twitter. In: *Proceedings of the 17th Association of Computing Machinery Special Interest Group on Knowledge Discovery and Data Mining International Conference on Knowledge Discovery and Data Mining*. New York, NY: Association of Computing Machinery; 2011: 430–438.

45. Burger JD, Henderson J, Kim G, Zarrella G. Discriminating gender on Twitter. In: *Proceedings of the Conference on Empirical Methods in Natural Language Processing*. Stroudsburg, PA: Association for Computational Linguistics; 2011: 1301–1309.

46. Ikeda K, Hattori G, Ono C, Asoh H, Higashino T. Twitter user profiling based on text and community mining for market analysis. *Knowl Base Syst*. 2013;51: 35–47.

47. Bruns A, Moe H. Structural layers of communication on Twitter. In: Weller K, Bruns A, Burgess J, Mahrt M, Puschmann C, eds. *Twitter and Society*. New York, NY: Peter Lang Publishing; 2014: 15–28.

48. Chary M, Genes N, McKenzie A, Manini AF. Leveraging social networks for toxicovigilance. *J Med Toxicol*. 2013;9(2):184–191.

49. Weller K, Bruns A, Burgess J, Mahrt M, Puschmann C. Twitter and society: an introduction. In: Weller K, Bruns A, Burgess J, Mahrt M, Puschmann C, eds. *Twitter and Society*. New York, NY: Peter Lang Publishing; 2014: xxix–xxxviii.

50. Rodrigues E, Assunção R, Pappa GL, Renno D, Meira W. Exploring multiple evidence to infer users' location in Twitter. *Neurocomputing*. 2016;171:30–38.

51. Tsou M-H. Research challenges and opportunities in mapping social media and Big Data. *Cartogr Geogr Inf Sci*. 2015;42(suppl 1):70–74.

52. Vo DT, Hai VT, Ock CY. Exploiting language models to classify events from Twitter. *Computational Intelligence and Neuroscience*. 2015;2015.

53. Centers for Disease Control and Prevention. National diabetes statistics report: estimates of diabetes and its burden in the United States, 2014. Atlanta, GA: US Department of Health and Human Services; 2014.

54. Nwankwo T, Yoon SS, Burt V, Gu Q. Hypertension among adults in the United States: National Health and Nutrition Examination Survey, 2011–2012. *NCHS Data Brief*. 2013;(133):1–8.

55. Mozaffarian D, Benjamin EJ, Go AS, et al. Heart disease and stroke statistics—2015 update: a report from the American Heart Association. *Circulation*. 2015;131(4): e29–e322.

56. Roger VL, Boerwinkle E, Crapo JD, et al. Strategic transformation of population studies: recommendations of the Working Group on Epidemiology and Population Sciences from the National Heart, Lung, and Blood Advisory Council and Board of External Experts. *Am J Epidemiol*. 2015;181(6):363–368.

57. Weeg C, Schwartz HA, Hill S, Merchant RM, Arango C, Ungar L. Using Twitter to measure public discussion of diseases: a case study. *JMIR Public Health Surveill*. 2015; 1(1):e6.

58. Lau AY, Gabarron E, Fernandez-Luque L, Armayones M. Social media in health—what are the safety concerns for health consumers? *Health Information Manage J*. 2012; 41(2):30–35.

59. Vayena E, Salathé M, Madoff LC, Brownstein JS. Ethical challenges of big data in public health. *PLOS Comput Biol*. 2015;11(2):e1003904.

60. Zimmer M. "But the data is already public": on the ethics of research in Facebook. *Ethics Inf Technol*. 2010; 12(4):313–325.