

Toward Real-Time Infoveillance of Twitter Health Messages

There is growing interest in conducting public health research using data from social media. In particular, Twitter “infoveillance” has demonstrated utility across health contexts. However, rigorous and reproducible methodologies for using Twitter data in public health are not yet well articulated, particularly those related to content analysis, which is a highly popular approach.

In 2014, we gathered an interdisciplinary team of health science researchers, computer scientists, and methodologists to begin implementing an open-source framework for real-time infoveillance of Twitter health messages (RITHM). Through this process, we documented common challenges and novel solutions to inform future work in real-time Twitter data collection and subsequent human coding.

The RITHM framework allows researchers and practitioners to use well-planned and reproducible processes in retrieving, storing, filtering, subsampling, and formatting data for health topics of interest. Further considerations for human coding of Twitter data include coder selection and training, data representation, codebook development and refinement, and monitoring coding accuracy and productivity. We illustrate methodological considerations through practical examples from formative work related to hookah tobacco smoking, and we reference essential methods literature related to understanding and using Twitter data. (*Am J Public Health*. Published online ahead of print June 21, 2018: e1–e6. doi:10.2105/AJPH.2018.304497)

Jason B. Colditz, MEd, Kar-Hai Chu, PhD, Sherry L. Emery, PhD, MBA, Chandler R. Larkin, BS, A. Everette James, JD, MBA, Joel Welling, PhD, and Brian A. Primack, MD, PhD

Roughly one quarter of adult Internet users in the United States use the Twitter social media platform, and as many as 95% of Twitter users make their posts publicly available.^{1,2} Because of user proliferation on this platform, there has been a steady increase in research using Twitter to monitor and characterize health communication and to stage and monitor public health interventions.^{3,4} Recent work has noted a need for public health practitioners to increase their social media engagement to combat misinformation, influence policy, and disseminate timely information.⁵ However, although social media “infoveillance” is a popular area of research,⁶ reproducible procedures for obtaining and working with these data are not well documented in the public health literature. Thus, there is a paucity of methodological guidance for public health professionals who wish to effectively use Twitter data. As a result, “practitioners have indicated a persistent gap between concept and routine implementation”^{7(p850)} of these data in public health.

Data from the Twitter platform provide insights into health topics such as tobacco use and cessation,^{4,8–10} cancer communication,^{11,12} mental health,^{13–15} vaccination,^{16–18} and public health policy.¹⁹ In addition to content and sentiment analysis, these data are useful for tracking diffusion of public health messaging.^{10,20} Although Twitter

data are not necessarily generalizable to broader populations outside the Twitter platform, they do present opportunities for infoveillance that complement well-established epidemiological approaches.²¹ Furthermore, the Twitter platform can be used in public health promotion by “implementing and disseminating critical prevention, screening, and treatment messages”^{22(p1)} in real time. As public health researchers and practitioners develop an increased interest in using Twitter to these ends, it will be increasingly important to have well-documented, reproducible processes for engaging with these data.

Methods of Twitter data analysis are inconsistent and not well documented across individual research studies.⁸ As data collection sources vary widely (e.g., collected directly from Twitter or obtained through various third-party providers), data quality is also a concern. Therefore, we sought to provide clarity on methodological considerations for obtaining and working with Twitter data among public health researchers

and practitioners. We provide examples of practical and theoretical challenges that we began working through in 2014 while developing a data framework with interdisciplinary collaboration among public health and policy researchers, computer scientists, and methodologists across research institutions. We provide practical considerations and concrete examples that relate to pilot research on hookah tobacco smoking, as this trend is both popular and has presented a variety of challenges while working with Twitter data. We refer to the broader research framework as real-time infoveillance of Twitter health messages (RITHM), which includes (1) open-source software for collecting and formatting Twitter data, and (2) procedures for maximizing the efficiency and effectiveness of subsequent human data coding.

DATA COLLECTION AND PROCESSING

Twitter messages (tweets) currently include up to 280 text

ABOUT THE AUTHORS

Jason B. Colditz, Kar-Hai Chu, Chandler R. Larkin, and Brian A. Primack are with the Center for Research on Media, Technology, and Health, University of Pittsburgh School of Medicine, Pittsburgh, PA. Sherry L. Emery is with NORC, University of Chicago, Chicago, IL. A. Everette James is with the Health Policy Institute, University of Pittsburgh Graduate School of Public Health, Pittsburgh. Joel Welling is with the Pittsburgh Supercomputing Center, Pittsburgh.

Correspondence should be sent to Jason Colditz, University of Pittsburgh School of Medicine, 230 McKee Place #600, Pittsburgh, PA 15213 (e-mail: colditzjb@pitt.edu). Reprints can be ordered at <http://www.ajph.org> by clicking the “Reprints” link.

This article was accepted April 14, 2018.
doi: 10.2105/AJPH.2018.304497

characters, embedded images, videos, and Web links. The text can also include emoji (in-text images depicting facial expressions, common objects, and activities). Twitter popularized hashtag use whereby individuals can preface a word or condensed phrase with the “#” symbol to link it to other uses of that hashtag within the platform, which helps users to exchange dialogues on particular topics.²³ Users can also interact via mentions (i.e., linking to a user name in a tweet), retweets (i.e., rebroadcasting someone else’s tweet), and quoting (i.e., embedding others’ tweets within a tweet).

Ethical Considerations

Twitter data present unique challenges related to the protection of individuals’ welfare and expectations of privacy.^{24–26} Although Twitter data are often relegated to nonhuman participant research because collected tweets are publicly available, basic standards for handling these data should always be observed to reduce the risk of identifying individual Twitter users in data sharing and dissemination. Foremost, in addition to censoring user names, it is important to not include direct quotations, images, or other characteristics that could later be used to identify observed individuals.

Additionally, as Twitter users may make tweets private or delete their accounts at any time, it is not appropriate to publicly archive or disseminate saturated data that contain detailed content or individual user characteristics. Rather, data sharing should be facilitated using desaturated data sets, including only tweet IDs, which can later be used to resaturate data that remain publicly available (i.e., deleted and private tweets are not retrieved). It is also

important to not share or disseminate particularly sensitive coded data (e.g., substance use, sexual behavior, health conditions) at the level of individual users or identifiable tweets. Appropriate data handling and sharing enhance the reproducibility of research while respecting Twitter users’ privacy and anonymity. Additional precautions also need to be considered, such as observing data use guidelines from the Twitter Developer Agreement and Policy (<https://developer.twitter.com/en/developer-terms/agreement-and-policy.html>), working within approved institutional review board protocols, and respecting broader ethical issues that apply to social media research.²⁶

Obtaining Real-Time Data

Twitter data can be obtained in real time directly from Twitter’s Public Streams Application Programming Interface (API). This is a versatile approach to data collection, but implementation can require an advanced grasp of software development.²⁷ Therefore, we developed an open-source RITHM streamer, which relies on basic template files to set data collection parameters, using the Python programming language for ease of use and versatility across operating systems. Twitter’s stream provides only prospective data and does not fill in missing data if the streamer stops running. This type of data collection process, although not requiring a great deal of computer processing power, is only as reliable as the hardware running it. An institutional or commercial server is ideal, although a desktop computer may suffice for small-scale projects.

Twitter’s filtered data stream includes all publicly available tweets that match a user-defined set of keyword parameters, up to a maximum of 1% of the total volume of tweets flowing at a certain time.²⁸ The *filtered* stream should not be confused with the *sample* stream, which delivers a pseudorandom 1% sample of all public tweets. To not exceed the 1% cap, it is important to consider the scope of keywords and other parameters (e.g., language, geocoordinates) used to filter the stream. Exceeding this limitation can result in nonrandom censoring of data,²⁸ which is a threat to generalizability. Similarly, it is important to understand and report on understood limitations of third-party data providers, if used, as these may rely on other methods for data capture, which would affect data randomization and generalizability and the validity of findings.

Scope of Collection

Data collection time frames greatly affect the generalizability of collected data and, conversely, present practical concerns about data overload. For example, in an early pilot study in 2014, we streamed data related to hookah tobacco smoking over the span of 2 distinct weekends, resulting in 43 092 tweets. As there are known variations in daily and weekly rhythms of Twitter use,²⁹ these data have limited generalizability outside the context of this data collection time frame. However, as the scope of data collection increases to bolster generalizability, the number of tweets obtained presents practical challenges for data storage. For example, on April 20, which is widely considered a “marijuana holiday,”³⁰ we consistently observe a dramatic spike in tweets

related to “smoking,” resulting in 7 to 8 gigabytes of data collected in a single day. Although this fluctuation in data collection is fairly predictable, other, less predictable events can occur when a particular topic or tweet goes viral and is heavily shared (e.g., a celebrity, political leader, or popular news source is the origin). Thus, data collection projects that are running long term or on a particularly broad or popular topic area should expect that peak data collection might be many magnitudes greater than that of a typical day.

Another methodological consideration that affects validity and data size relates to selecting appropriate keywords to target data collection.³¹ Having too few keywords will result in missing valuable data, whereas including too many is likely to clutter the data with irrelevant content. For example, in the hookah-related work we used 5 primary keywords (“hookah,” “hooka,” “shisha,” “sheesha,” and “narghile”). We chose these keywords to account for common terminology and misspellings. We purposefully excluded non-specific terms such as “water-pipe” because in preliminary searches they resulted in a substantial amount of false-positive references to cannabis paraphernalia. Before final keywords are chosen for inclusion or exclusion, initial searches should be performed and documented using established criteria for assessing search performance, such as precision (specificity) and recall (sensitivity).^{31,32} This process helps to demonstrate generalizability and validity within broader Twitter data and can help to acknowledge data collection limitations when reporting search procedures and findings.

Another technical limitation of the Twitter stream is that

individual keywords do not match at the level of a complete phrase. Instead, the streaming API combines consecutive words with a logical AND operator. For example, the search string “smoke sessions” (i.e., social gatherings, such as at a hookah lounge) matches a number of tweets about US Attorney General Jeff Sessions that contain the word “smoke” in other contexts. The Twitter API does not support a NOT operator to exclude tweets that contain words, such as “Jeff” in this example. This limitation can be overcome programmatically. However, integrating additional data searching and formatting procedures into an active streamer process can delay data capture. If the delay compounds over many tweets, the API will eventually disconnect and lose the backflow of data. To avoid data loss, primary data collection should implement a “dump-and-go” approach with additional data searching and formatting procedures happening in a separate process (e.g., RITHM parser).

Formatting

The Twitter API delivers raw data in JavaScript Object Notation (JSON) format, where individual tweets have unequal numbers of nested data fields (i.e., not in row and column format). Within these data are additional fields related to extraneous display parameters, such as image size, as well as redundant data (e.g., hashtags appear in multiple fields). An important step in data parsing is to reformat the data to a comma-separated spreadsheet format. In-text formatting such as commas and carriage returns then need to be replaced with placeholders such as “_comma_” or “_return_” so

that the spreadsheet retains proper formatting. Similarly, as the raw data are text based, emoji are delivered as Unicode strings that need to be converted to a human-readable format (e.g., “\u2764” converts to “_heart_” for the “❤️” emoji).

As emoji are displayed and interpreted differently across diverse computer and telephone platforms,³³ this text-based approach indicates the presence of particular emoji without making assumptions about display parameters. Although new emoji are continually added to the Unicode Standard (<http://www.unicode.org/standard>), an emoji dictionary must be regularly updated to recognize new objects. Unrecognized objects should remain intact as Unicode strings so that additional procedures can be used in situations in which translation is needed (e.g., Arabic script, ideographic characters). Overall, it is important to consider that tweets contain a variety of text and nontext characters that add substantial clarity, and these nuances should be preserved and portrayed in a human-readable format as much as possible.

Subsampling and Generalization

When human coders assess data, selecting the number of tweets to code presents additional balancing of feasibility (e.g., how much effort is available for coding procedures) versus generalizability (e.g., what proportion of tweets is representative). Subsampling needs to be addressed on a project-by-project basis. For codebook development, it may be appropriate to use a randomized subsample of all available tweets. For analysis, more complex subsampling procedures may be warranted to account for

the relative prevalence of tweets over time. In this case, a “keep every n th tweet” approach or a randomized subsample stratified by observed tweet prevalence may be preferable.

Other approaches include stratifying by keyword prevalence or reshuffling subsamples until a particular optimization parameter is met. For example, assessing ranked correlations between popular hashtags in the full sample versus a subsample can help to demonstrate content validity of a sampling strategy.⁹ When selecting a data reduction approach, it is important to consider that this does not bolster data generalizability or validity to circumstances outside Twitter. Rather, subsampling can make human coding more feasible while demonstrating the generalizability of the coded data within a broader data set.

HUMAN CODING

A recent systematic review of studies using Twitter data identified content analysis as the most common methodological approach.³ However, approaches to content coding and reporting methods vary widely.⁸ So another aim of our work was to develop reproducible methods for human coding of tweets. Through our pilot work, we synthesized formative feedback about this process from coders who were involved. Human coding is inherently time consuming and not necessarily suitable for implementation in a real-time process. However, certain considerations and approaches can ease coder burden and greatly expedite the process.

Coder Considerations

Coders require attention to detail, an understanding of the

overarching research goals for the coding task, and appropriate background experience to accurately interpret the content.³⁴ Coders who are Twitter users may be more adept at interpreting idiosyncratic linguistic conventions of tweets. For example, “hookah is bae” would be more easily interpretable by someone familiar with “bae” as an acronym for “before anyone else,” a highly valued person or thing. Someone unfamiliar with this convention could easily interpret “bae” as a typographical error for “bad” and misjudge the message sentiment. This is 1 of many examples of common shorthand used to communicate complex ideas within the brevity of a tweet. When possible, we find it helpful to work with coders who are familiar with current “netspeak” conventions and have some personal experience using Twitter or similar platforms.

Another aspect to consider is familiarity with the health topic under investigation. Before coding hookah-related tweets, coders were first immersed in common contexts of hookah use, which included viewing pictures of hookah lounges and videos of individuals participating in hookah smoking sessions. Coders also reviewed and discussed hookah-related tweets from the Twitter Web site before formal coding. This type of training can be foundational to reducing coding discrepancies early on. However, in some circumstances it may also be beneficial for coders to be naïve to subject matter (e.g., coding related to general public perceptions, coding within a grounded theory framework). In general, clearly describing coder expectations, expertise, and training methods are integral to ensuring the reproducibility of coding work.

Portraying Tweets

It is important to consider tradeoffs between presenting tweets as text only for more efficient coding versus viewing content directly on the Twitter Web site for more contextual nuances (e.g., images, emoji). We eventually moved to a middling approach by which coders could evaluate tweets using basic textual information and had the option to use the Twitter Web site as a complement if the text was unclear. In hookah-related data, this led us to discover a context-dependent emoji that was particularly salient. The `_dash_` emoji depicts a cartoonish dust cloud left behind when someone dashes away. In the context of hookah, this emoji was consistently used to signify a puff of smoke, which it also resembles.³⁵ Depending on the topic and observational variables of interest, imagery and emoji may be particularly useful for coders to observe in graphical format.

After trying various approaches, a simple spreadsheet seemed the most straightforward

way to portray tweets to coders. Specifically, a coding spreadsheet should include (at a minimum) the tweet ID, the tweet text, a link to the online tweet, and additional columns for entering coded variables. This limits the technical complexities and irreproducibility that might come with developing custom databases or data entry portals. It also reduces coder burden and ensures portability of data sets across analytic platforms.

CODING TWEET CONTENT

Although coding definitions may differ among studies, broad categories of codes such as relevance, sentiment, and theme are commonly employed.⁸ Initial coding for relevance helps to narrow the number of tweets needing additional coding and also allows the calculation of search precision and recall estimates.³¹ Sentiment can be evaluated in a variety of ways.⁸ For example, a statement such as “so happy to finally quit smoking”

conveys positive sentiment with regard to general affect but negative sentiment toward smoking. This can be a point of coder confusion. Alternately, framing sentiment as optimistic versus pessimistic might be more useful for approaching content such as “quitting smoking is too hard.” Clearly defining the intention and scope of sentiment codes allows improved interrater reliability (Figure 1), clearer interpretation of results, and more expedient coding. Finally, thematic coding accounts for contextual factors. These are often dichotomous categories, such as health related (Table 1). Examining sentiment across well-defined thematic codes (deductive or inductive derived codes) allows more nuanced interpretations of content and deeper qualitative synthesis.⁹

Hookah Smoking Codebook

For a practical example from the hookah data, our codebook was first broadly framed with deductive codes of relevance and sentiment. Inductive codes such

as commercial and health related were refined later, from a grounded codebook development process. Decisions need to be made early on about tradeoffs for inductive or deductive codebook development approaches and adapting from an existing coding framework when possible.^{34,36} This initial hookah codebook was drafted over 3 rounds of coders evaluating 200 randomly selected tweets, mapping potential constructs, identifying exemplar tweets, and coming to consensus on salient codes and preliminary definitions. Coding definitions were further refined through independent double-coding, meetings to adjudicate coder disagreements, and then updates to the coding framework. Using snapshots of independently double-coded data before adjudication of disagreements, we assessed interrater reliability using the Cohen κ coefficient (Figure 1). This occurred at semiregular intervals, ranging between 200 and 500 tweets, depending on coding progress and timing of adjudication meetings. To establish validity and generalizability of results, it is important to follow well-established content coding methods, including independent double-coding and reporting interrater reliability coefficients along commonly used metrics.^{34,37}

As this particular codebook evolved, we added several dozen example keywords and tweets to help add nuance to coding definitions. However, interrater reliability did not improve in a linear fashion over initial coding iterations (Figure 1). Coders also expressed frustration at codebook definitions becoming overly complex and more arduous to interpret. After reflecting on the trajectory of codebook development, we simplified the coding definitions, removed highly

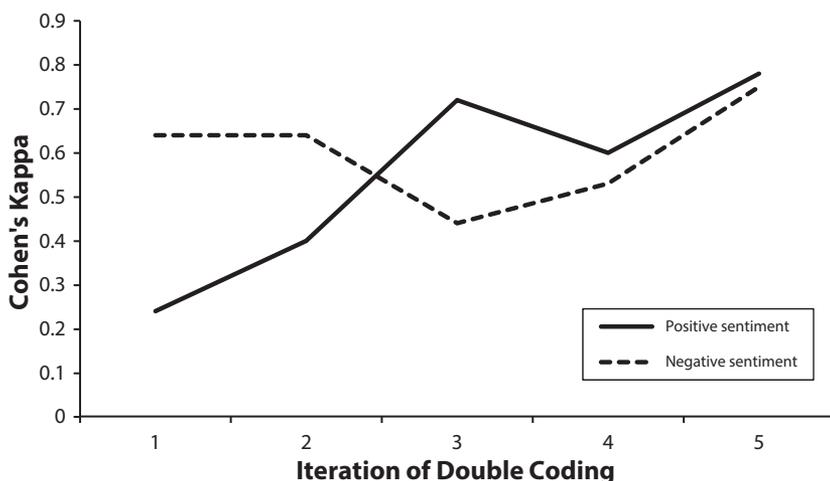


FIGURE 1—Interrater Agreement for Assessing Sentiment During Codebook Development Procedures: 2015–2016

TABLE 1—Final Coding Definitions for Tweets Relevant to Hookah: 2016

Variable	Definition	Examples
Pro-hookah (formerly positive sentiment)	The tweeter demonstrates positive attitude toward hookah	Smoking hookah or wants/plans to smoke hookah Recently used hookah Mentions a song/music about hookah References sex or romance and hookah
Anti-hookah (formerly negative sentiment)	The tweeter demonstrates negative attitude toward hookah	Health harms and other negative effects of hookah Hookah smoking is unattractive, uncool, or disgusting Quitting smoking hookah Doesn't smoke hookah or doesn't want to try smoking hookah Prefers to use a different substance, such as marijuana, as opposed to hookah smoking Stereotyping hookah culture or hookah users
Context: commercial	The tweet is commercial in nature	An advertisement for hookah products or vendors A promotion of hookah at bars/lounges/events
Context: health related	The tweet contains content that is health related	Health harms and other negative effects of hookah Hookah is not unhealthy or is healthier than something else (e.g., cigarettes) Mentions things like tar, nicotine Mentions health policy (e.g., FDA regulations)

Note. FDA = US Food and Drug Administration.

nuanced examples, and further clarified sentiment as “pro-hookah” or “anti-hookah” instead of “positive” or “negative” more broadly. The final codebook more closely reflected a content coding approach that we had previously seen in the literature,³⁸ which would have ultimately been a better starting point to save time and effort in codebook development. With the simplified codebook (Table 1), interrater reliability was greatly improved in the fifth coding iteration (Figure 1). In general, we found that fewer codes with shorter and more targeted codebook definitions (e.g., pro-hookah vs positive sentiment) resulted in both faster and more accurate tweet coding. Furthermore, it is important to carefully review the literature for past content analysis work before engaging in extensive codebook development expeditions.

Coder Logistics

Depending on coder availability, it can be difficult to

schedule meetings for coders to adjudicate coding disagreements and refine codebook definitions. In light of these challenges, we developed a process of asynchronous adjudication. Through this process, coders had the opportunity to review coding disagreements with the option of changing their own coding assessments if they felt that they had overlooked a clear circumstance that the other coder had noticed. Coders also recorded notes of any emerging considerations to refine coding definitions. When in-person meetings occurred, both coders would discuss a reduced number of disagreements, and they used their notes to frame discussions on codebook refinement. When appropriate, this process can substantially expedite codebook development and adjudication.

A coding log can further help monitor overall coding progress, track milestones, and set achievable goals. In our pilot work, this also brought coders' attention to common distractions that interrupted coding. On the basis of feedback in 48 log entries of 2

experienced coders, coding took a median of 9 seconds per tweet and proceeded for a median of 76 tweets before a distraction occurred. In open-ended descriptions of the process, coders described distractions such as “zoning out” or “overthinking” some tweets. Coders' reflections on the process indicated that the added structure of having a coding log was extrinsically motivating and helped them maintain focus. Long periods of coding can be cognitively challenging, and coders preferred to alternate between coding tweets and other tasks to reduce cognitive demand.

IMPLICATIONS AND RECOMMENDATIONS

This narrative serves as an overview of some basic challenges and associated strategies for working with Twitter data that are collected in real time and analyzed by human coders. We covered methodological considerations that have downstream

impacts on generalizability and validity of findings and practical approaches for facilitating coding. Further methodological discussion is warranted on more specialized approaches to analysis (e.g., machine learning, network analysis, geospatial inference). Additional ethical and methodological considerations are also needed on the evaluation of Twitter user profiles, which present an important and underused source of research insights.³ We hope that considerations presented will inform continued research and methodological critiques.

Future systematic reviews might evaluate the extent to which recently published Twitter research implements rigorous and reproducible approaches at this basic level. This will continue to inform best practices in future research. Overall, we hope that this narrative (1) serves as a useful primer for practical considerations related to undertaking Twitter research, (2) guides researchers toward more consistently reporting on the

methodological rationales and decisions framing particular approaches, and (3) provides a framework for thoughtfully assessing the methodological rigor of studies using Twitter data. Additionally, the RITHM streamer and parser code are open-source and currently available through <https://github.com/CRMTH/RITHM>. **AJPH**

CONTRIBUTORS

J. B. Colditz, K-H. Chu, and S. L. Emery contributed to developing procedures and practical considerations for undertaking this work. J. B. Colditz, C. R. Larkin, and J. Welling developed the RITHM software infrastructure. A. E. James and B. A. Primack developed the initial concept and institutional infrastructure for this project. All authors contributed to the conceptualization and design of the overarching framework and essay development.

ACKNOWLEDGMENTS

This work was supported by the Pittsburgh Supercomputing Center (grant TG-DBS160002), the National Cancer Institute (grant R01CA225773), and the University of Pittsburgh Health Policy Institute (internal funding).

We thank Denys Lau for feedback on the initial conceptualization of this essay and Michelle Woods for editorial assistance. We also appreciate the efforts of research assistants who diligently coded and provided critical feedback on the coding process: Erica Barrett, Beth Hoffman, Christine Stanley, Maharsi Naidu, Daria Williams, and Tabitha Yates.

HUMAN PARTICIPANT PROTECTION

This work was approved by the University of Pittsburgh institutional review board.

REFERENCES

- 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. 2016. Available at: <http://www.pewinternet.org/2016/11/11/social-media-update-2016>. Accessed May 22, 2018.
- Liu Y, Kliman-Silver C, Mislove A. The tweets they are a-changin': evolution of Twitter users and behavior. In: *Proceedings of the 8th International AAAI Conference on Weblogs and Social Media*. Palo Alto, CA: Association for the Advancement of Artificial Intelligence; 2014:305–314.
- Sinnenberg L, Buttenheim AM, Padrez K, Mancheno C, Ungar L, Merchant RM. Twitter as a tool for health research:

a systematic review. *Am J Public Health*. 2017;107(1):e1–e8.

- Lakon CM, Pechmann C, Wang C, Pan L, Delucchi K, Prochaska JJ. Mapping engagement in Twitter-based support networks for adult smoking cessation. *Am J Public Health*. 2016;106(8):1374–1380.
- Breland JY, Quintiliani LM, Schneider KL, May CN, Pagoto S. Social media as a tool to increase the impact of public health research. *Am J Public Health*. 2017; 107(12):1890–1891.
- Eysenbach G. Infodemiology and infoveillance: tracking online health information and cyberbehavior for public health. *Am J Prev Med*. 2011;40(5 suppl 2): S154–S158.
- Burkom HS. Evolution of public health surveillance: status and recommendations. *Am J Public Health*. 2017;107(6):848–850.
- Lienemann BA, Unger JB, Cruz TB, Chu K-H. Methods for coding tobacco-related Twitter data: a systematic review. *J Med Internet Res*. 2017;19(3):e91.
- Colditz JB, Welling J, Smith NA, James AE, Primack BA. World vaping day: contextualizing vaping culture in online social media using a mixed methods approach. *J Mix Methods Res*. 2017;1–20. [Published online April 9, 2017]
- Chung JE. A smoking cessation campaign on Twitter: understanding the use of Twitter and identifying major players in a health campaign. *J Health Commun*. 2016;21(5):517–526.
- Koskan A, Klasko L, Davis SN, et al. Use and taxonomy of social media in cancer-related research: a systematic review. *Am J Public Health*. 2014;104(7): e20–e37.
- Pemmaraju N, Thompson MA, Mesa RA, Desai T. Analysis of the use and impact of Twitter during American Society of Clinical Oncology annual meetings from 2011 to 2016: focus on advanced metrics and user trends. *J Oncol Pract*. 2017; 13(7):e623–e631.
- McClellan C, Ali MM, Mutter R, Kroutil L, Landwehr J. Using social media to monitor mental health discussions—evidence from Twitter. *J Am Med Inform Assoc*. 2017;24(3):496–502.
- Braithwaite SR, Giraud-Carrier C, West J, Barnes MD, Hanson CL. Validating machine learning algorithms for Twitter data against established measures of suicidality. *JMIR Ment Health*. 2016; 3(2):e21.
- Berry N, Lobban F, Belousov M, Emsley R, Nenadic G, Bucci S. #Why-WeTweetMH: understanding why people use Twitter to discuss mental health problems. *J Med Internet Res*. 2017;19(4): e107.
- Chakraborty P, Colditz JB, Silvestre AJ, Friedman MR, Bogen KW, Primack BA. Observation of public sentiment

toward human papillomavirus vaccination on Twitter. *Cogent Med*. 2017;4(1):1–10.

- Kang GJ, Ewing-Nelson SR, Mackey L, et al. Semantic network analysis of vaccine sentiment in online social media. *Vaccine*. 2017;35(29):3621–3638.
- Dunn AG, Surian D, Leask J, Dey A, Mandl KD, Coiera E. Mapping information exposure on social media to explain differences in HPV vaccine coverage in the United States. *Vaccine*. 2017; 35(23):3033–3040.
- Feng M, Pierce JP, Szczypka G, Vera L, Emery S. Twitter analysis of California's failed campaign to raise the state's tobacco tax by popular vote in 2012. *Tob Control*. 2017;26(4):434–439.
- Carpenter JS, Laine T, Harrison B, et al. Topical, geospatial, and temporal diffusion of the 2015 North American Menopause Society position statement on nonhormonal management of vasomotor symptoms. *Menopause*. 2017;24(40): 1154–1159.
- Allem JP, Escobedo P, Chu KH, Soto DW, Cruz TB, Unger JB. Campaigns and counter campaigns: reactions on Twitter to e-cigarette education. *Tob Control*. 2017;26(2):226–229.
- Xu S, Markson C, Costello KL, Xing CY, Demissie K, Llanos AA. Leveraging social media to promote public health knowledge: example of cancer awareness via Twitter. *JMIR Public Health Surveill*. 2016;2(1):e17.
- Zappavigna M. Searchable talk: the linguistic functions of hashtags. *Soc Semiotics*. 2015;25(3):274–291.
- Zimmer M, Proferes NJ. A topology of Twitter research: disciplines, methods, and ethics. *Aslib J Inf Manag*. 2014;66(3): 250–261.
- Williams ML, Burnap P, Sloan L. Towards an ethical framework for publishing Twitter data in social research: taking into account users' views, online context and algorithmic estimation. *Sociology*. 2017;51(6):1149–1168.
- Hunter RF, Gough A, O'Kane N, et al. Ethical issues in social media research for public health. *Am J Public Health*. 2018; 108(3):343–348.
- Kumar S, Morstatter F, Liu H. *Twitter Data Analytics*. New York, NY: Springer; 2014.
- Morstatter F, Liu H. Discovering, assessing, and mitigating data bias in social media. *Online Soc Networks Media*. 2017;1: 1–13.
- Liang H, Fu K. Testing propositions derived from Twitter studies: generalization and replication in computational social science. *PLoS One*. 2015;10(8): e0134270.
- Bravo AJ, Pearson MR, Conner BT, Parnes JE. Is 4/20 an event-specific marijuana holiday? A daily diary

investigation of marijuana use and consequences among college students. *J Stud Alcohol Drugs*. 2017;78(1):134–139.

- Kim Y, Huang J, Emery S. Garbage in, garbage out: data collection, quality assessment and reporting standards for social media data use in health research, infodemiology and digital disease detection. *J Med Internet Res*. 2016;18(2):e41.
- Stryker JE, Wray RJ, Hornik RC, Yanovitzky I. Validation of database search terms for content analysis: the case of cancer news coverage. *J Mass Commun Q*. 2006;83(2):413–430.
- Miller H, Thebault-Spieker J, Chang S, Johnson I, Terveen L, Hecht B. "Blissfully happy" or "ready to fight": varying interpretations of emoji. In: *Proceedings of the Tenth International Conference on Web and Social Media*. Menlo Park, CA: Association for the Advancement of Artificial Intelligence Press; 2016:259–268.
- Krippendorff K. Recording/Coding. In: *Content Analysis: An Introduction to Its Methodology*. Los Angeles, CA: Sage; 2013: 126–149.
- Grant A, O'Mahoney H. Portrayal of waterpipe (shisha, hookah, nargile) smoking on Twitter: a qualitative exploration. *Public Health*. 2016;140:128–135.
- Elo S, Kyngäs H. The qualitative content analysis process. *J Adv Nurs*. 2008; 62(1):107–115.
- Krippendorff K. Reliability in content analysis: some common misconceptions and recommendations. *Hum Commun Res*. 2004;30(3):411–433.
- Krauss MJ, Sowles SJ, Moreno MA, et al. Hookah-related Twitter chatter: a content analysis. *Prev Chronic Dis*. 2015; 12:E21.