Using Social Media for Research and Public Health Surveillance

P.I. Eke

Division of Adult and Community Health, National Center for Chronic Disease Prevention and Health Promotion, Centers for Disease Control and Prevention, 1600 Clifton Rd., Atlanta, GA 30333, USA; peke@cdc.gov

The article in this issue of *JDR* by Heaivilin and colleagues with the title 'Public Health Surveillance of Dental Pain *via* Twitter" (Heaivilin *et al.*, 2011) introduces a potential new data source for dental surveillance and research, namely, publicly available information from the social network medium "Twitter". The authors present a novel idea and approach in using publicly available Twitter data to assess dental pain experiences. Undoubtedly, monitoring episodes of dental pain, including the impact of the pain and actions taken to relieve pain, is a worthwhile objective for dental public health and has indeed been assessed in previous population-based surveys such as in the National Health and Nutrition Survey (NHANES) and National Health Interview Surveys (NHIS) (Beltrán-Aguilar *et al.*, 2005; NIDCR/CDC DRC, 2011). This perspective provides a brief critical assessment of the use of Twitter for public health surveillance and research.

Public health surveillance is the ongoing systematic collection, analysis, and interpretation of health data from defined populations for use in planning, implementing, and evaluating public health programs (Thacker and Berkelman, 1988). The most important attributes of public health surveillance systems include simplicity, flexibility, and acceptability of the data collection instruments, as well as sensitivity, positive predictive value, representativeness, and timeliness of the data collected (Romaguera *et al.*, 2000). It can be argued that tools such as Twitter do possess some of these attributes. Notably, Twitter data are available publically, and the data are relatively simple to access, extract, and analyze, as exemplified by the study by Heaivilin's group (Heaivilin *et al.*, 2011). Furthermore, tweets are reported in real time by millions of real persons from across several continents and are communicated *via* a variety of simple and easy-to-use formats, which are increasingly accessible in most populations.

However, surveillance systems are more deliberate and specific in their purpose, and their methods are standardized, reproducible, and geared toward collecting valid and reliable data over time from a strictly defined target population. In addition, surveillance systems are sufficiently flexible to accommodate new objectives and changes in definitions or procedures to improve the quality of data collected. In contrast, Twitter users start or join conversations about a specific topic (*e.g.*, dental pain) by exchanging instant messages, each up to 140 characters long, with other users. Thus, tweeters are not responding to questions in

a standard questionnaire whose content can be modified by the investigator to capture responses to other more specific subjects of interest, *e.g.*, "Are you having dental pain now?" Also, Twitter is not branded as a survey tool and therefore, "Twitterers" are not knowingly participating in a study and are unaware that their tweets will be used to assess health status, both of which factors can influence their responses. The lack of both anonymity and consent to participate in a survey and challenges for maintaining mandatory confidentiality of personal information raise concerns that will have to be addressed before tweets can be used for surveillance.

Consequently, investigators using Twitter will rely on proxies or the extraction of certain phrases singly or in combination (such as "toothache" and "dental pain") to identify persons with dental pain. This can be problematic, because this process does not account for the context in which these terms were used and may result in low predictive value positive for detecting persons with dental pain. In the study by Heaivilin *et al.* (2011), the terms "toothache" and "dental pain" were used collectively and interchangeably, even though they may represent different subjective experiences of dental pain, possibly including a broader spectrum of diagnostic entities and etiologies (*e.g.*, acute irreversible pulpitis, dental abscess, caries, periodontal conditions, trauma, cracked tooth syndrome, and myofacial pain). This ambiguity in diagnosis and etiology threatens the validity and reliability of the estimates of dental pain frequency and the potential value of the information for guiding public health programs that deliver evidence-based dental advice to an at-risk population.

More importantly, a basic attribute of surveillance systems is that the data collected represent health events that occurred in a particular population classified by person, place, and time. This system attribute, lacking in data derived from Twitter, represents the greatest weakness of using tweets for surveillance. External validity concerns include that Twitter data are skewed toward active users who are often young adults, well-educated white females, and persons who more likely live in higher income households (Lorica, 2010; Webster, 2010). Any results based on Twitter data exclude people who do not use Twitter, who are likely to be the most vulnerable in populations, and who are often unwilling to share their health experiences publicly. Intuitively, it can be speculated that persons who are ill, elderly, in discomfort, or disabled would be less likely to tweet, as would those who are illiterate or not 'computer savvy.' Consequently, the Twitter population is skewed toward a subset of the world's population that excludes key segments of the general population in most countries in great need of the services offered by dental public health programs.

In the example of the study by the Heaivilin team (Heaivilin *et al.*, 2011), the only personal characteristics reported were for sex and location. In the context of surveillance, the information on location has limited usefulness, because the geopolitical or administrative unit for analysis is unclear. This has important implications for determining population statistics and measures of prevalence and incidence, which require well-defined denominator measures in the population. For example, the 43.3% of Twitter users who reported seeing a dentist can be interpreted as the average of dental visits by users spanning 4 continents and other unknown locations. Thus, the sensitivity of this information relative to a local community is unknown, and therefore, the usefulness and implications for public health monitoring and action are unknown. Similarly, the quality of the information obtained from

Twitter is questionable, because information extracted from Twitter, such as sex (based on name, user name, and photo posted) and location (based on user-identified location or the time zone), are not direct measures and have not been validated, which introduces a potential source of information bias into the system. The more important basic information on age and socio-economic characteristics that are critical in public health surveillance and research for identifying at-risk populations and for providing dental advice over the lifespan is not provided.

In conclusion, Heavilin *et al.* have illustrated a novel approach to using data from Twitter to explore public opinions about a dental health issue (dental pain experiences) and to understand challenges to be overcome in planning research studies and in conducting routine public health surveillance in well-defined populations of Twitter users. Researchers in the health communication field have used archives of Twitter messages to examine the public's response to public health emergencies, such as to make recommendations for the use of mammography in screening for breast cancer (Squiers *et al.*, 2011). Also, the extensive reach of Twitter is currently being used successfully in public health to distribute health information to the segments of the public who access Twitter. For example, the Centers for Disease Control and Prevention (CDC) currently use Twitter as part of a larger communication and social media strategy to disseminate accurate health messages quickly and widely (CDC, 2009, 2010). At this time, however, there are major limitations and challenges to be overcome before Twitter and its data products can be used for routine public health surveillance and research in the general populations in a particular public health jurisdiction at the local, state, or national level.

Acknowledgments

References

Beltrán-Aguilar ED, Barker LK, Canto MT, Dye BA, Gooch BF, Griffin SO, et al. Surveillance for dental caries, dental sealants, tooth retention, edentulism, and enamel fluorosis—United States, 1988-1994 and 1999-2002. MMWR Surveill Summ. 2005; 54:1–43.

Centers for Disease Control and Prevention. [6/7/2011] H1N1 Web and social media metrics; cumulative data report; April 22, 2009 – December 31, 2009. 2009. at: http://www.cdc.gov/metrics/campaigns/reports/h1n1-cumulative_report_01-31-10.pdf

Centers for Disease Control and Prevention. [6/7/2011] CDC eHealth Metrics Dashboard; Twitter; Annual summary (2010). 2010. at: http://www.cdc.gov/metrics/socialmedia/micro-blogs.html

Heaivilin N, Gerbert B, Page JE, Gibbs JL. Public health surveillance of dental pain *via* Twitter. J Dent Res. 2011; 90:1047–1051. [PubMed: 21768306]

Lorica, B. [6/7/2011] Twitter by the numbers. 2010. at: http://radar.oreilly.com/2010/04/twitter-by-the-numbers.html

National Institute of Dental and Craniofacial Research (NIDCR)/Centers for Disease Control and Prevention. [6/7/2011] Dental, Oral, and Craniofacial Data Resource Center (DRC). 2011. at: http://drc.hhs.gov

Romaguera, RA., German, RR., Klaucke, DN. Evaluating public heath surveillance. In: Teutsch, SM., Churchill, RE., editors. Principles and practice of public health surveillance. 2. New York, NY: Oxford University Press; 2000. p. 176-193.

Squiers LB, Holden DJ, Dolina SE, Kim AE, Bann CM, Renaud JM. The public's response to the U.S. Preventive Services Task Force's 2009 recommendations on mammography screening. Am J Prev Med. 2011; 40:497–504. [PubMed: 21496748]

Thacker SB, Berkelman RL. Public health surveillance in the United States. Epidemiol Rev. 1988; 10:164–190. [6/7/2011] at: http://epirev.oxfordjournals.org/content/10/1/164.long. [PubMed: 3066626]

Webster, T. [6/7/2011] Twitter usage in America: 2010. 2010. at: http://www.edisonresearch.com/twitter_usage_2010.php