

Commentary

Using Social Media and Internet Data for Public Health Surveillance: The Importance of Talking

DAVID M. HARTLEY

Georgetown University Medical Center

It doesn't have to be like this. Our greatest hopes could become reality in the future with the technology at our disposal. The possibilities are unbounded. All we need to do is make sure we keep talking.

—Stephen Hawking¹

DESPITE PROGRESS IN PUBLIC HEALTH AND THE biomedical sciences, infection has yet to be vanquished: vaccine-preventable diseases continue to be transmitted; pandemics occur; previously unknown pathogens emerge; contaminated foods and food products are traded and consumed; and the specter of a post-antibiotic era looms ever larger. Bioterrorism is, and will remain, a danger. Infectious disease is both a national and an international security issue and represents an important threat to human health and well-being.

In order to confront these and related threats, detailed data regarding the global ebb and flow of disease are needed. Over many decades, surveillance methods (often termed “indicator-based” methods) have been developed and refined to provide disciplined, standardized approaches to acquiring and recording important information. More recently, ubiquitous and unstandardized data collected from the Internet have been used to gain insight into emerging disease events. Although this approach—known as “Internet-based biosurveillance,” “digital disease detection,” or, more simply, “event-based” surveillance—has been described and

Address correspondence to: David M. Hartley, Department of Microbiology and Immunology, 2115 Wisconsin Ave NW, Suite #603, Washington, DC 20057 (email: david@no-infection.net).

The Milbank Quarterly, Vol. 92, No. 1, 2014 (pp. 34-39)
© 2014 Milbank Memorial Fund. Published by Wiley Periodicals Inc.

analyzed in the literature,²⁻⁴ systematic reviews of the field have been few.

It is this intellectual gap that makes the article by Edward Velasco and his coworkers in this issue of the *Quarterly* so valuable and timely. Velasco and his colleagues systematically searched for and reviewed more than 20 years of published studies of event-based systems and approaches, providing a much-needed perspective on both research in the field and several important issues. After selecting relevant peer-reviewed studies to include in their analysis, they extracted the attributes of 13 different event-based systems. They then defined 15 different descriptive attributes that capture the principal facets of event-based systems, including the languages and types of diseases systems covered, the methods by which each system produces its output, and the types of users that each system attracts. Such metrics are necessary for comparing and contrasting different approaches to event-based surveillance. Readers should bear in mind that the properties and lifetimes of these systems are dynamic, as is the Internet itself, and that technologies and methodologies change rapidly, allowing systems to improve and evolve over weeks or months. Accordingly, one of the key contributions of Velasco and colleagues' study is the set of metrics they propose, in which event-based systems can be tracked over time in order to quantitatively understand how much event-based biosurveillance has changed and continues to change.

Velasco and colleagues seek to provide a basis for public health agencies incorporating event-based methods into existing, comprehensive surveillance programs, and they cite user confidence in this approach as an important step in this process. Their review of the literature, however, uncovered no event-based systems that were regularly incorporated into national programs for surveillance during their study period (1990-2011). Moreover, they found no comprehensive evaluations showing whether or not these systems had been deployed during real-time health events.

Although this evidence may be lacking in the peer-review literature included in their study, there is evidence that several systems are utilized, to varying extents, by national and international public health organizations. At the international level, for example, the World Health Organization (WHO) uses the Canadian-based Global Public Health Intelligence Network in its global alert

and response activities.⁵ The European Centre for Disease Prevention and Control utilizes the MedISys system (<http://medusa.jrc.it/medisys/homeedition/en/home.html>),⁴ and a recent study described the evaluation of several event-based systems by international public health professionals.⁶ At a national level, the US Centers for Disease Control and Prevention (US CDC) utilize event-based data,⁷ and at the local level, a social media–monitoring program known as Foodborne Chicago is being used to monitor foodborne diseases.^{8–9} Because the information from the WHO,⁵ the US CDC,⁷ and Foodborne Chicago⁸ are web pages or newspaper stories published after the study period,⁹ rather than peer-reviewed studies produced during the study period, Velasco and colleagues did not include them. This is less a criticism than an illustration of how quickly event-based data are evolving and of why such information is not necessarily wholly contained in the research literature. Consequently, it will be critical for future studies to include public media and non-peer-reviewed sources in their assessments of event-based data systems.

Of course, a broader question is that regardless of how many public health workers are currently using these systems, what is preventing them from being utilized more broadly and effectively? Here, the approach used in a recent work by Barboza and colleagues is instructive.⁶ They asked respondents to rate, on a uniform scale, the usability and relative strengths of several event-based systems. The results highlighted the complementarity of different systems and demonstrated the value of using multiple systems to produce the most robust results from the event-based approach. In combination, that study and the work by Valasco and colleagues underscore the importance of consulting stakeholders in the design and refinement of event-based surveillance systems. Accordingly, an assessment of stakeholder engagement would be a useful metric to include in future systematic reviews.

Velasco and colleagues discuss the limitations of event-based systems as well, such as (1) information is not always moderated by professionals or interpreted for relevance before it is disseminated to interested surveillance epidemiologists; (2) there is no standardized system for updates, often resulting in too much information; (3) algorithms and statistical baselines are not well developed; and (4) new information related to

health events or probable cases is not always disseminated in the most efficient way. These limitations point to two vital issues.

First, different users have different needs. Some need to see everything reported by event-based surveillance systems (ie, although they are not concerned about specificity, they are concerned about sensitivity), whereas other users may demand low false-alarm rates (ie, specificity is important to their needs). Put another way, some users are more interested in early warnings of threats, so they need to examine all indications of an emerging event. Others, however, are more interested in the situational awareness of identified threats. Thus, interpreting Valasco and colleagues' findings in the context of diverse users' needs is paramount.^{10,11}

Second, users must be involved in the design and revision of event-based systems in order to address their specific requirements. This point is central to achieving not only a wider use of event-based surveillance but also its more effective use. If event-based surveillance is to be broadly recognized as a timely modality available to government and public health officials, health care workers, and the public and private sectors, this approach must be refined and strengthened in accordance with methodological, engineering, and user support perspectives.³

One of the most promising new event-based surveillance methods is the use of social media in what is known as "participatory epidemiology." An example is Flu Near You (<https://flunearyou.org/>), a system in which any individual 13 years of age or older and living in the United States or Canada can register to complete weekly surveys regarding influenza-like illnesses near them. The information on the site is available to public health officials, researchers, disaster-planning organizations, and the general public, with a mobile application available in addition to a Web interface. Such an approach makes it easy for nonspecialists to contribute, in an open and transparent way, data that may provide a valuable addition to indicator-based surveillance (eg, the U.S. Outpatient Influenza-like Illness Surveillance Network [<http://www.cdc.gov/flu/weekly/overview.htm>]). The use of mobile applications to collect information, as well as to view and access it in the field, represents an important trend in event-based surveillance.

Finally, for both practical applications and user confidence, determining more precisely whether these systems can improve the early detection and rapid response to infectious outbreaks is important.^{12,13}

One promising example of this trend was recently reported by Chunara and coworkers on the use of Internet-based social and news media to enable the estimation of epidemiological patterns early during the 2010 outbreak of cholera in Haiti.¹⁴ Their research team was able to estimate the basic reproductive ratio (R_0) in that outbreak, a feat difficult even under normal circumstances using carefully collected epidemiologic data in the field.

For all these reasons, so nicely articulated in the Velasco article, it is safe to state that novel sources of event-driven epidemiological data—along with their accurate use and analysis—will play an even greater role in epidemics and pandemics not yet experienced or even imagined.

References

1. Baker M. *Translation and Conflict: A Narrative Account*. London: Routledge; 2006:150.
2. Brownstein JS, Freifeld CC, Madoff LC. Digital disease detection—harnessing the Web for public health surveillance. *N Engl J Med*. 2009;360:2153-2155.
3. Hartley D, Nelson N, Walters R, et al. Landscape of international event-based biosurveillance. *Emerg Health Threats J*. 2010;3:e3. <http://www.eht-journal.net/index.php/ehtj/article/view/7096>. Accessed October 30, 2013.
4. Hartley D, Nelson RN, Arthur P, et al. An overview of Internet biosurveillance. *Clin Micro Infect*. 2013;19:1006-1013.
5. World Health Organization. Epidemic intelligence—systematic event detection. <http://www.who.int/csr/alertresponse/epidemicintelligence/en/>. Accessed October 30, 2013.
6. Barboza P, Vaillant L, Mawudeku A, et al. Evaluation of epidemic intelligence systems integrated in the early alerting and reporting project for the detection of A/H5N1 influenza events. *PLoS ONE*. 2013;8:e57252. <http://www.plosone.org/article/info%3Adoi%2F10.1371%2Fjournal.pone.0057252>. Accessed October 30, 2013.
7. US Centers for Disease Control and Prevention. Building USG interagency collaboration through global health engagement. <http://www.cdc.gov/washington/EGlobalHealthEditions/eGlobalHealth0408.htm>. Accessed October 30, 2013.
8. Smart Chicago Collaborative. Foodborne Chicago. <http://foodborne.smartchicagoapps.org/>. Accessed October 30, 2013.

9. Eng M. Food-poisoning tweets get city follow-up. Health authorities seek out sickened Chicagoans, ask them to report restaurants. *Chicago Tribune*. August 13, 2013.
10. World Health Organization. WHO Technical consultation on event-based surveillance; March 19-21, 2013; Lyon, France. http://www.episouthnetwork.org/sites/default/files/meeting_report_ebs_march_2013_final.pdf. Accessed October 30, 2013.
11. Corley CD, Lancaster MJ, Brigantic RT, et al. Assessing the continuum of event-based biosurveillance through an operational lens. *Biosecur Bioterror*. 2012;10:131-141.
12. Chan EH, Brewer TF, Madoff LC, et al. Global capacity for emerging infectious disease detection. *Proc Natl Acad Sci USA*. 2010;107:21701-21706.
13. Tsai FJ, Tseng E, Chan CC, Tamashiro H, Motamed S, Rouge-mont AC. Is the reporting timeliness gap for avian flu and H1N1 outbreaks in global health surveillance systems associated with country transparency? *Global Health*. 2013;9(14). <http://www.globalizationandhealth.com/content/9/1/14>. Accessed October 30, 2013.
14. 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;86:39-45.