



HHS Public Access

Author manuscript

N Engl J Med. Author manuscript; available in PMC 2016 May 19.

Published in final edited form as:

N Engl J Med. 2013 August 1; 369(5): 401–404. doi:10.1056/NEJMp1307752.

Influenza A (H7N9) and the Importance of Digital Epidemiology

Marcel Salathé, Ph.D., Clark C. Freifeld, M.S., Sumiko R. Mekaru, D.V.M., Anna F. Tomasulo, M.P.H., and John S. Brownstein, Ph.D.

Center for Infectious Disease Dynamics, Department of Biology, Pennsylvania State University, University Park (M.S.); and the Informatics Program, Boston Children's Hospital (C.C.F., S.R.M., A.F.T., J.S.B.), the Department of Biomedical Engineering, Boston University College of Engineering (C.C.F.), the Department of Epidemiology, Boston University School of Public Health (S.R.M.), and the Department of Pediatrics, Harvard Medical School (J.S.B.) — all in Boston.

On March 31, 2013, Chinese health officials notified the World Health Organization of three cases of human infection with novel influenza A (H7N9). Since then, 132 people have been infected, 37 of them fatally (see figure, Panel A). To date, there is no evidence of ongoing human-to-human transmission. However, a number of characteristics of this virus are cause for heightened attention.

First, the human population has not been exposed on a large scale to hemagglutinin 7 (H7) and neuraminidase 9 (N9) antigens in recent times and therefore most likely lacks immunity against this strain. Second, initial analyses of viral genome sequences suggest signs of adaptation to mammals — such as the ability to attach to respiratory tissue and to replicate at mammalian body temperatures¹ — which could facilitate human-to-human transmission. Third, because the virus has low pathogenicity in birds, its presence is difficult to detect in bird flocks by symptomatic surveillance alone. Testing of more than 20,000 people with influenza-like illness in China revealed only six cases of H7N9 infection, suggesting that there are not widespread mild cases of H7N9.²

Public health officials on the ground continue to focus substantial resources on assessing and mitigating the pandemic potential of this virus strain. Although these efforts are critical for understanding the evolving public health situation, since there are limited resources available, intelligence for assessing the threat must come from a wide range of data sources. Though relatively new, digital disease surveillance is an increasingly powerful tool that complements traditional approaches.³

In this and other outbreaks, digital disease surveillance has supplemented the critical laboratory studies and work in the trenches by public health officials and epidemiologists, by leveraging widespread use of the Internet, mobile phones, and social media.³ Many of these added insights come from the general population, whose access to technology enables rapid information flow. In 2013, there are 6.8 billion cellphone subscribers; 2.7 billion people are online; and by the end of the year, there will be more than 2 billion mobile broadband subscriptions worldwide. A large percentage of the online population publicly shares

information on social media services: in both the United States and China, for example, more than half the population with access to the Internet uses social media services.

Digital data can be used in at least four ways for studying infectious-disease dynamics. First, they can be used for early detection of disease outbreaks. This capacity was illustrated most recently in China, when a hospital employee uploaded an image of the medical record of a patient with H7N9 infection to Sina Weibo, a popular Chinese social network similar to Twitter. The post was promptly deleted, but it appears to have accelerated the government's acknowledgment of four new cases (see figure, Panel B). More generally, because digital surveillance is not limited by the hierarchies of traditional public health infrastructure, geographic communication barriers, and geopolitical obstacles, it has improved the timeliness of outbreak detection substantially in recent years.⁴

Second, these data can be used to continuously monitor disease levels. With proper filtering by automated systems (see the *Journal's* H7N9 HealthMap tracking system [<http://healthmap.org/h7n9>]), analyst-driven systems (e.g., the Global Public Health Intelligence Network of Canada), vigilant journalists on Twitter (e.g., Crawford Kilian [[@Crof](#)] and Helen Branswell [[@HelenBranswell](#)]), and crowd-sourced systems (e.g., FluTrackers and ProMED-mail), informal data sources such as news media, e-mail lists, blogs, and social media can complement formal public health surveillance by offering real-time clues to disease dynamics. Internet-based surveillance systems provided important early epidemic intelligence during the 2003 outbreak of severe acute respiratory syndrome (SARS) and the 2009 H1N1 influenza pandemic, enhancing transparency by rapidly publicizing outbreak information.³

Third, Internet-based data from social media can be used to assess disease-relevant health-related behaviors and sentiments relevant to disease control. During the H1N1 pandemic, sentiments about vaccination extracted from Twitter were shown to correlate well geographically with subsequent vaccination coverage throughout the United States.⁵ Such analyses could provide important information to aid in planning and in the distribution of limited resources, as well as improving public health communications efforts.

Fourth, these data provide researchers with an additional method for examining the period before an outbreak came to light. Despite international agreement that transparency is critical during an outbreak, accusations of delayed reporting are common and can be difficult to dispel. Time-series analysis of the volume of influenza-related searches on the Chinese Web search engine Baidu shows a low level of activity in the months leading up to the first announced H7N9 cases, which suggests that widespread unreported outbreaks were not festering before the announcement.

Though digital epidemiology as an enabler of disease surveillance across political, cultural, and linguistic borders carries many advantages, it's not without its challenges. The dynamics of information spread are inherently different from the dynamics of disease spread. In the days immediately following the first reports of a few cases of a new disease, the volume of news reports and social media posts typically spikes dramatically (see figure, Panel A), because during that period most information is new and potentially relevant and therefore of

high interest to the public. After some time, information saturation sets in, and public interest wanes, even as the number of new cases continues to rise. For newly emerging diseases with initially few cases, the social media data are typically dominated by news reports rather than first-person accounts of symptoms by sick individuals — a situation that differs markedly from that during recurring epidemics such as seasonal influenza. In the early phase of an outbreak, this effect makes it challenging for digital epidemiology to provide intelligence for early detection of cases of disease. Both human assessment and computational algorithmic intelligence are required to meet the challenge of extracting information from data sets that are both extremely large and noisy.

In addition, information retrieval itself can be difficult. Consider Sina Weibo, which can be a particularly challenging platform for harvesting information. Posts are censored at a reportedly rapid rate, with 5% of deletions happening within 8 minutes, and 30% within 30 minutes. Because of concerns about censors, some Sina Weibo users rely on metaphors (see http://chinadigitaltimes.net/space/Introduction_to_the_Grass-Mud_Horse_Lexicon for an interesting glossary of terms used in the past), making it challenging for outsiders to follow what they're talking about. For example, in Chinese, the eponymous “grass-mud horse” sounds like an obscene phrase and indirectly references the Communist party. The phrase has evolved to mean an Internet-savvy person who dislikes and seeks to circumvent government censorship.

Despite these challenges, the recent outbreaks of H7N9 influenza and the Middle East respiratory syndrome coronavirus (MERS-CoV) infections illustrate the strengths of digital disease surveillance. In the case of H7N9, such surveillance has enhanced transparency and helped public health officials to understand the outbreak more fully. Although information was sparse in the MERS-CoV outbreak, digital disease surveillance proved its usefulness: the initial MERS-CoV case reports came to light through ProMED-mail (www.promedmail.org/direct.php?id=20120920.1302733). Since the SARS outbreak, the world has seen substantial progress in transparency and rapid reporting. The extent of these advancements varies, but overall, digital disease surveillance is providing the global health community with tools supporting faster response and deeper understanding of emerging public health threats.

Health officials are aware of the catastrophic potential of pandemics. The potential for widespread infections with MERS-CoV outside of Saudi Arabia and the potential reemergence of H7N9 during next year's influenza season demand that digital disease surveillance be a part of the response.

references

1. Uyeki TM, Cox NJ. Global concerns regarding novel influenza A (H7N9) virus infections. *N Engl J Med.* 2013; 368:1862–4. [PubMed: 23577629]
2. Xu, C.; Havers, F.; Wang, L., et al. Monitoring avian influenza A(H7N9) virus through national influenza-like illness surveillance, China. Centers for Disease Control and Prevention; Atlanta: 2013. (http://wwwnc.cdc.gov/eid/article/19/8/13-0662_article.htm)
3. Brownstein JS, Freifeld CC, Madoff LC. Digital disease detection — harnessing the Web for public health surveillance. *N Engl J Med.* 2009; 360:2153–7. [PubMed: 19423867]

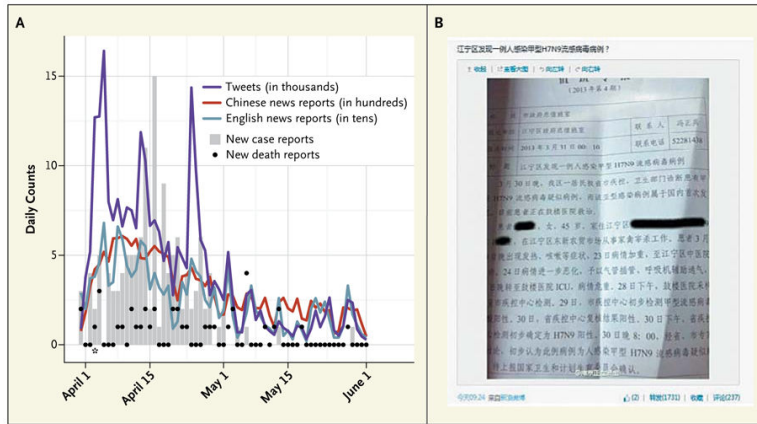
4. Chan EH, Brewer TF, Madoff LC, et al. Global capacity for emerging infectious disease detection. *Proc Natl Acad Sci U S A*. 2010; 107:21701–6. [PubMed: 21115835]
5. Salathé M, Khandelwal S. Assessing vaccination sentiments with online social media: implications for infectious disease dynamics and control. *PLoS Comput Biol*. 2011; 7(10):e1002199. [PubMed: 22022249]

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript



Social Media Activity, News Volume, and Reports of Cases and Deaths Associated with H7N9 Influenza, March 31 through June 1, 2013.

Panel A shows daily counts of tweets containing the keyword “H7N9,” Chinese and English news reports (from the *Journal*s and HealthMap’s H7N9 tracking system), and new reports of cases and deaths. Panel B shows the medical record of a patient with H7N9 influenza, posted on the Chinese microblogging website Sina Weibo on April 3, 2013. The asterisk in Panel A indicates the date when the patient record shown in Panel B was posted on Sina Weibo.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript