

FEATURE

PUBLIC HEALTH

Can Twitter predict disease outbreaks?

Connie St Louis and Gozde Zorlu examine how public health experts are beginning to exploit the power of social media

Connie St Louis *director, science journalism MA*¹, Gozde Zorlu *freelance journalist*²

¹City University, London, UK ; ²London

In March 2011 the most powerful earthquake and tsunami in Japan's history caused horrifying devastation on the country's northeastern coast. Along with a massive loss of life, the entire infrastructure of the region was destroyed: buildings were crushed and telephone lines were down. However, the mobile internet was still available, and resourceful doctors decided to use Twitter to inform chronically ill patients where they could obtain essential medicines. In a letter to the *Lancet* Yuichi Tamura and Keiichi Fukuda, cardiologists at Keio University School of Medicine in Tokyo, wrote: "We were able to notify displaced patients via Twitter on where to acquire medications. These 'tweets' immediately spread through patients' networks, and consequently most could attend to their essential treatments."¹

Today the success of Twitter continues unabated, with over 500 million accounts and more than half of active users signing in every day. And it's not just Twitter. The use of social media has quadrupled in the past five years: Facebook has more than 800 million active users, and WordPress, one of the most popular blogging platforms, holds over 15 million blogs.

Use of social media by doctors to communicate with patients raises a multitude of ethical conundrums, particularly verification of identity. But the medical potential of this untapped source of data is beginning to be recognised. Infectious disease experts and computer scientists are working together to use this open data to improve disease surveillance.

Quicker detection

Public health agencies rely on traditional methods of surveillance to monitor outbreaks of disease. These include collection of diagnostic information from doctors and laboratory reporting of test results. Although this way of gathering data is very accurate, it can take a long time to identify new outbreaks and orchestrate a response. And time is critical when trying to prevent rapid spread of a disease.

What has caught the attention of infectious disease experts is the growing number of informal sources of information that can provide a much faster picture of outbreaks.

Digital surveillance platforms such as HealthMap (www.healthmap.org) and BioCaster (<http://born.nii.ac.jp>) use software programs that visit websites on the internet to search and extract information—a process known as trawling and scraping. They monitor news and social media sites, including blogs, to pick up clues about emerging public health threats, but the information is less accurate and needs to be verified.

According to the first systematic review of studies looking at the use of data mined from social networking sites for prediction and monitoring disease, "Social media represents a new frontier in disease surveillance."²

Quicker detection means more time to prepare resources. An analysis of three million tweets between May and December 2009, by Patty Kostkova and her colleagues at City eHealth Research Centre, City University, London, showed that the 2009 H1N1 flu outbreak could have been identified on Twitter one week before it emerged in official records from general practitioner reports.³ The results of their study are corroborated by another project called Flusurvey (<http://flusurvey.org.uk>), set up during the flu pandemic to collect data from the general public on influenza rates across England.

"The speed is useful," said Ken Eames, who runs the website at the London School of Hygiene and Tropical Medicine, adding that "an extra week or two can be massively important in preparing a response."

The Flusurvey project, part of a European initiative to monitor influenza trends, collects data from over 2000 volunteers who log on every week to report any flu-like symptoms. It provides a useful addition to the traditional methods of surveillance because most people with flu do not see their general practitioner.

Data from this website are published in the Health Protection Agency's weekly influenza update. Eames hopes that, along

with the other Flusurvey projects being carried out across Europe, they may be able to collaborate with the European Centre for Disease Control and Prevention (ECDC). “What we would like is for the data we use to feed into the ECDC surveillance platform, in the same sort of way that the data from the UK Flusurvey are made available to the HPA.”

But Eames is cautious about replacing traditional based surveillance systems with such websites.

“All of these tools are useful, but I don’t think any one of them does it perfectly,” he said. Getting all the data streams and tools together is the really big job that will have to happen at some point.”

The next stage of development of Flusurvey is to extend the number of volunteers reporting to the website through a mobile phone application. This should increase the accuracy of the information being collected.

John Brownstein, cofounder of HealthMap and an epidemiologist at Harvard Medical School and Children’s Hospital Boston, believes that social media are set to revolutionise epidemiology. “It’s a new field [and] a lot of work still needs to be done, but there’s a huge amount of promise. Informal web based data are not bound the way that official data are, and they can provide a vastly different image of current public health. Digital methods of disease detection should continue to be explored so that we learn how to correct for biases and have information that allows us to prevent, anticipate, and respond to epidemics,” he said.

During the Olympics in London this summer, potential outbreaks will be monitored by a collaborative project between HealthMap, Bio.Diaspora (which uses information on air travel to predict disease), and the Health Protection Agency. It could be the first time a surveillance system predicts an outbreak in real time.

Disaster response

Obtaining information directly from the public through informal sources is particularly valuable when local outbreaks are not covered by traditional surveillance systems. In many countries surveillance systems are not as robust as in the UK because of social, economic, or political constraints, and natural disasters can also disrupt collection of data.

After the Haitian earthquake, researchers used HealthMap’s automated surveillance system to chart the cholera outbreak. HealthMap looks at trends in the volume of reporting in informal sources, such as Twitter and news media, as well as collecting some data from official reports. Gathering data by this route made information on the distribution of cholera cases available two weeks before official sources released it.⁴

While internet access may be less widely available in developing countries, there is no shortage of mobile phones. In 2009, there were 3.2 billion subscriptions in the developing world. So for the millions of people who are displaced by natural disasters every year, phones are an excellent way to access information.

Linus Bengtsson’s research group at the Karolinska Institute in Stockholm has made use of the ubiquity of mobile phones. After the 2010 earthquake and cholera outbreak in Haiti they successfully tracked population movements using anonymous SIM card data from mobile phone providers. He is now working with mobile phone companies to mine data from phones in disasters to track population movements for relief agencies and to be able to detect disease outbreaks.⁵

Official uses

“Healthcare agencies have been actively using media surveillance of big events for a number of years. However, incorporating social media systems into full operation is premature,” said Kostkova.

There have been various collaborations between HealthMap and healthcare agencies such as the US Centers for Disease Control and Prevention (CDC), ECDC, and the World Health Organization. However, more work, such as the collaborative project for the Olympic Games needs to be done to prove its value and how it would fit in with traditional mechanisms.

Although the ECDC uses information from specialist blogs, such as FluTracker, it does not yet use data from Facebook and Twitter on a daily basis.

“The tools are not mature enough. However, for some specific situations, they provide an added value,” said Thomas Mollet from the centre’s surveillance and response support unit.

Mollet outlined two examples of how the ECDC uses Facebook and Twitter. In March 2011, there was a cluster of sudden deaths among tourists in Chiang Mai, Thailand. Although the ECDC was unable to confirm the cause using Facebook, officials were able to use the site to ascertain where the tourists were from, their location, and details about their trip. Over the next few weeks, the World Health Organization and Thai authorities undertook an investigation and it seems that the origin was a toxin discovered in their hotel, although there is still no strong evidence to support this.

Last year, there was a suspected outbreak of pneumonic plague in Burma, where for political reasons information from local media was not thorough enough for surveillance. Although monitoring of Twitter picked up rumours of plague, two doctors tweeted that these were cases of diarrhoea. Three weeks later, WHO ruled out plague.

Mollet remains optimistic: “There are a lot of new initiatives. It’s amazing how many people are interested in this field, and I’m really convinced that in the coming years, we will use social media on a daily basis. Facebook and Twitter are not able to confirm or rule out an outbreak, but they contribute to an investigation,” said Mollet.

Taha Kass-Hout, director for information science at CDC in Atlanta is also optimistic about the future: “The general consensus is that we have to pay more attention to social media.”

He said the main benefits are shortening the length of time it takes to detect outbreaks to improve responses and allow healthcare agencies faster communication with the public.

Another unexpected byproduct that Kass-Hout has found, is that social media have made it easier for healthcare agencies and governments to share data with their citizens.

Referring to the severe acute respiratory syndrome (SARS) outbreak, Kass-Hout said: “We could have done a lot more if we found out a lot earlier, especially if there were more transparent ways of communicating it.”

But he emphasised that more work needs to be done to prove the value of social media as a prediction and tracking tool. As more and more information becomes available the background noise of these sites increases exponentially and with it, rumours and half truths. More models are needed to filter and validate the data from these informal sites.

The challenges posed by the veracity of social media information remain central whether it is used for gathering disease intelligence or urgent doctor-patient communication. Kostkova believes these challenges can be overcome through working

partnerships between specialist teams in surveillance units and the research community: “Close collaboration between public health agencies and ehealth researchers is essential to develop robust social media mining tools validated on large surveillance datasets to substantially enhance the quality and speed of epidemic intelligence.”

Competing interests: The authors have completed the ICMJE unified disclosure form at www.icmje.org/coi_disclosure.pdf (available on request from the corresponding author) and declare no support from any organisation for the submitted work; no financial relationships with any organisation that might have an interest in the submitted work in the previous three years; and no other relationships or activities that could appear to have influenced the submitted work.

Provenance and peer review: Commissioned; not externally peer reviewed.

- 1 Tamura Y, Fukuda K. Earthquake in Japan. *Lancet* 2011;377:1652.
- 2 Guy S, Ratzki-Leewing A, Bahati R, Gwadyri-Sridhar F. Social media: a systematic review to understand the evidence and application in infodemiology. *eHealth* 2011, 21-23 November 2011, Malaga, Spain. http://proceedings.dev.icstweb.eu/ehealth2011/papers/Ehealthpaper30-camerareadyversion_FINAL19Oct2011.pdf
- 3 Szomszor M, Kostkova P, de Quincey E. #swineflu: Twitter predicts swine flu outbreak in 2009. 3rd International ICST Conference on Electronic Healthcare for the 21st Century (eHealth2010), 13-15 December 2010, Casablanca, Morocco. www.mendeley.com/research/swineflu-twitter-predicts-swine-flu-outbreak-2009/#page-1
- 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;86:39-45.
- 5 Bengtsson L, Lu X, Thorson A, Garfield R, von Schreeb J. Improved response to disasters and outbreaks by tracking population movements with mobile phone network data: a post-earthquake geospatial study in Haiti. *PLoS Med* 2011;8:e1001083.

Cite this as: *BMJ* 2012;344:e2353

© BMJ Publishing Group Ltd 2012