

# Scientists examine use of big data for infectious disease surveillance

November 14 2016

---

Big data derived from electronic health records, social media, the internet and other digital sources have the potential to provide more timely and detailed information on infectious disease threats or outbreaks than traditional surveillance methods. A team of scientists led by the National Institutes of Health reviewed the growing body of research on the subject and has published its analyses in a [special issue](#) of *The Journal of Infectious Diseases*.

Traditional infectious disease surveillance—typically based on laboratory tests and other data collected by public health institutions—is the gold standard. But, the authors note it can have time lags, is expensive to produce, and typically lacks the local resolution needed for accurate monitoring. Further, it can be cost-prohibitive in low-income countries. In contrast, big data streams from internet queries, for example, are available in real time and can track disease activity locally, but have their own biases. Hybrid tools that combine traditional surveillance and big [data sets](#) may provide a way forward, the scientists suggest, serving to complement, rather than replace, existing methods.

"The ultimate goal is to be able to forecast the size, peak or trajectory of an outbreak weeks or months in advance in order to better respond to infectious disease threats. Integrating big data in surveillance is a first step toward this long-term goal," says Cecile Viboud, Ph.D., co-editor of the supplement and a senior scientist at the NIH's Fogarty International Center. "Now that we have demonstrated proof of concept by comparing data sets in high-income countries, we can examine these models in low-

resource settings where traditional surveillance is sparse."

Experts in epidemiology, computer science and modeling collaborated on the supplement's 10 articles. They report on the opportunities and challenges associated with three types of data: medical encounter files, such as records from healthcare facilities and insurance claim forms; crowdsourced data collected from volunteers who self-report symptoms in near real time; and data generated by the use of [social media](#), the internet and mobile phones, which may include self-reporting of health, behavior and travel information to help elucidate disease transmission.

But big data's potential must be tempered with caution, the authors say. Non-traditional data streams may lack key demographic identifiers such as age and sex, or provide information that underrepresents infants, children, the elderly and developing countries. Social media outlets may not be stable sources of data, as they can disappear if there is a loss of interest or financing. Most importantly, any novel data stream must be validated against established infectious disease surveillance data and systems, the authors said.

Each article features a promising example of the use of big data to monitor and model infectious diseases activity:

- In the United States, researchers found what they describe as "excellent alignment" between medical insurance claim data for flu-like illnesses and proven influenza activity reported by the Centers for Disease Control and Prevention.
- A European surveillance system that began collecting crowdsourced data on influenza as part of a research project is now considered an adjunct to existing surveillance activities. Influenzanet uses standardized online surveys to gather information from volunteers who self-report their symptoms on a weekly basis. A number of European Union member states are

now using the tool and expanding it to include Zika, salmonella and other diseases.

- An online platform, ResistanceOpen, was developed by U.S. and Canadian scientists to monitor antibiotic resistance at the regional level. The site takes advantage of publicly available, online data from community healthcare institutions as well as regional, national and international bodies. An analysis showed online information compared favorably with traditional reporting systems in the two countries.
- Multiple studies have looked at social media and internet health forums for information on drug use and to detect adverse drug reactions. While there are technical and ethical challenges, the authors suggest internet search logs and social media posts can provide information more quickly than traditional physician-based reporting systems.
- In a comparison of the relatively new field of epidemic forecasting to the better-established one for weather forecasting, the authors note the former is much more difficult given that there is less observational data for disease, and because human behavior has the potential to rapidly alter the course of an epidemic.
- An examination of spatial data—including from insurance claims and social media posts—shows their potential for filling geographical information gaps but also presents technical, practical and privacy challenges that must be addressed.
- With appropriate safeguards to ensure anonymity, call data records from mobile phones may provide "an unprecedented opportunity" to determine how travel affects disease transmission. Studies of malaria and rubella in Kenya showed call data improved the understanding of the spatial transmission of those diseases.
- Online news articles and health bulletins from public health agencies were manually extracted and modeled to elucidate

transmission patterns for two recent outbreaks—the Ebola epidemic in West Africa and a Middle East Respiratory Syndrome outbreak in South Korea. Internet findings were in line with traditional data, providing a proof of concept that this approach can be generalized and automatized to a variety of online sources and generate information on disease transmission.

- Researchers also describe the benefits of a novel, publicly available epidemic simulation data management system, called epiDMS, which provides storage and indexing services for large data simulation sets, as well as search functionality and data analysis to aid decision makers during healthcare emergencies.

While the new hybrid models that combine traditional and digital disease surveillance methods show promise, the scientists agree there is still an overall scarcity of reliable surveillance information, especially compared to other fields such as climatology, where the data sets are huge. "To be able to produce accurate forecasts, we need better observational data that we just don't have in [infectious diseases](#)," notes Professor Shweta Bansal of Georgetown University, a co-editor of the supplement. "There's a magnitude of difference between what we need and what we have, so our hope is that big data will help us fill this gap."

Multi-disciplinary initiatives such as the NIH-led Big Data to Knowledge program will be instrumental in expanding the use of [big data](#) in research, as noted in the supplement.

Provided by National Institutes of Health

Citation: Scientists examine use of big data for infectious disease surveillance (2016, November 14) retrieved 26 April 2024 from <https://medicalxpress.com/news/2016-11-scientists-big-infectious-disease-surveillance.html>

This document is subject to copyright. Apart from any fair dealing for the purpose of private study or research, no part may be reproduced without the written permission. The content is provided for information purposes only.