Role of big data in the early detection of Ebola and other emerging infectious diseases

The lack of adequate disease surveillance systems in Ebola-affected areas has both reduced the ability to respond locally and has increased global risk. There is a need to improve disease surveillance in vulnerable regions, and digital surveillance could present a viable approach.

Digital surveillance seeks to gain knowledge of public health issues through the analysis of data in the digital domain (such as internet search metrics, Twitter posts, or online news stories), the distribution of these data, and patterns of access. It has already shown some promise.[1] In 2002, the Global Public Health Intelligence Network, a news-feed aggregator developed by the Public Health Agency of Canada, provided the first alert of SARS (more than 2 months before publication by WHO) and prompted the confirmation of an emerging disease event by the Chinese Government.[2] A more recently developed system, HealthMap,[3] is currently applying a similar data-aggregation approach to monitor the evolving Ebola outbreak; HealthMap identified news stories reporting a strange fever in Guinea on March 14, 2014—9 days before the release of official case information for the ongoing Ebola outbreak.

Currently, the most comprehensively investigated digital surveillance approach is based on monitoring of internet search metrics.[4] These systems work on the premise that people who contract a disease are likely to seek information on their condition on the internet and an estimate of disease in the community can be produced by monitoring the frequency of specific searches. Overall, results for this approach have been promising; the scope of research has, however, been limited to a small number of diseases, particularly influenza, and has mainly focused on industrialised countries.

We used Google Trends to assess the volume and location of Google searches for "ebola" between Jan 1, 2014, and Oct 27, 2014. Search volume data were downloaded from Google Trends on Oct 28, 2014. Internet search volume increased markedly from affected regions over the course of the epidemic, and most Google searches during 2014 for "ebola" originated from the regions most affected (Liberia, Guinea, and Sierra Leone; figure). Furthermore, search
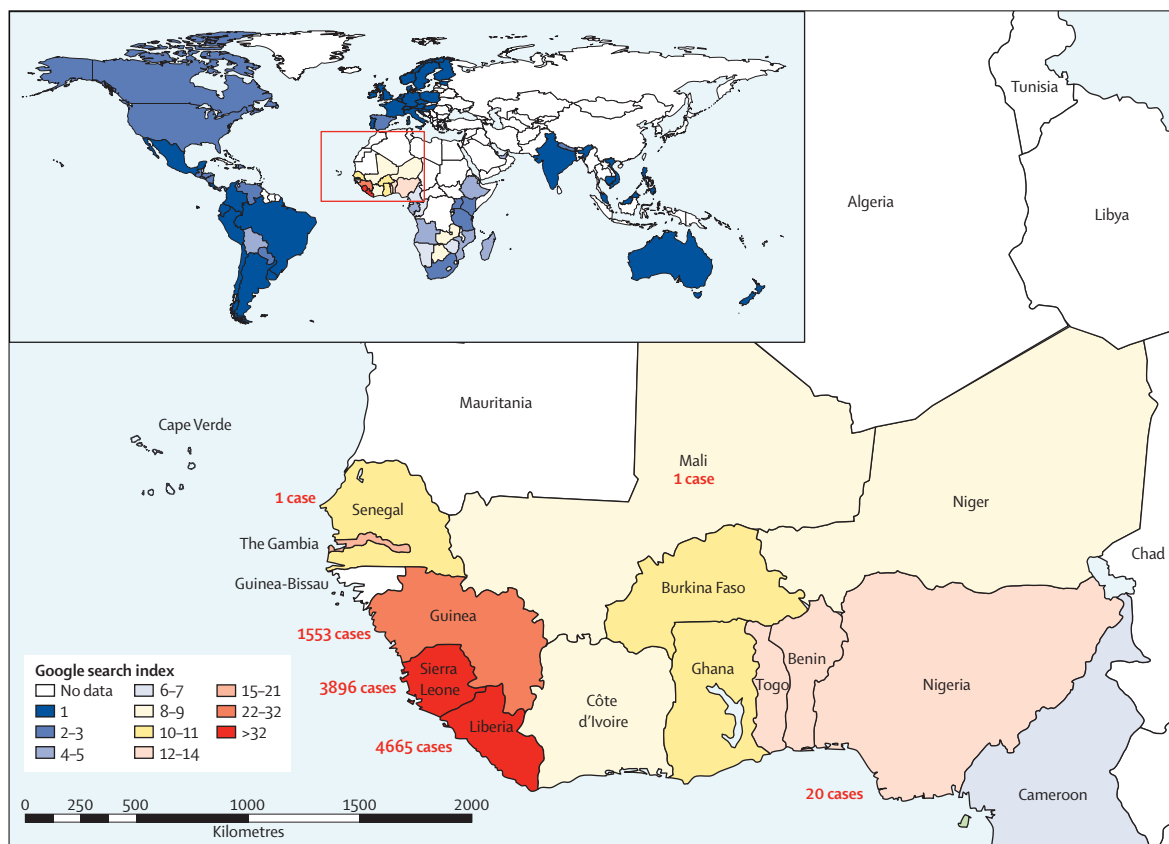
Figure: Relative internet search interest for "ebola" during 2014
The region with the highest search volume (Liberia) is indexed to 100 and other regional values are scaled accordingly. Total case numbers (as of Oct 25, 2014) are indicated beside the affected countries.[5]

frequency in these countries was highly correlated with epidemic curves (appendix). Pairwise correlations, by Spearman's Rho, between weekly national case numbers[5] and Google Trends search frequencies for the term "ebola" were 0·54 for Guinea, 0·70 for Liberia, 0·68 for Sierra Leone (all $p<0·001$; one-tailed).

There is clearly some noise in these data and this is not unexpected considering analyses only used a single search term and given the significant public and media interest in the outbreak. Also, whether digital surveillance systems might have facilitated earlier detection or reduced the effect of the ongoing Ebola epidemic remains moot. Overall, however, these results are promising and, generally, support the development of internet-based surveillance systems for developing countries.

Shifting patterns of health-seeking behaviour, the digitisation of society, and increased technology uptake present an opportunity to address emerging infectious disease events. Internet use in developing countries is high and continues to grow; as such, sufficient infrastructure exists on which to develop digital infectious disease surveillance and early warning systems in many of these regions.

There will be challenges to developing digital surveillance systems for regions that are not currently covered sufficiently by traditional surveillance systems. The potential impact for these regions, however, extends beyond the local scale. These systems will have global relevance and could contribute to improved global health security.

*Gabriel J Milinovich,
Ricardo J Soares Magalhães,
Wenbiao Hu
gabriel.milinovich@qut.edu.au

School of Public Health and Social Work, Queensland University of Technology, Kelvin Grove, 4056 QLD, Australia (GJM, WH); and School of Veterinary Science, University of Queensland, Gatton, 4343 QLD, Australia (RJSM)

1   Anema A, Kluberg S, Wilson K, et al. Digital surveillance for enhanced detection and response to outbreaks. *Lancet Infect Dis* 2014; **14:** 1035–37.

2   Mykhalovskiy E, Weir L. The Global Public Health Intelligence Network and early warning outbreak detection: a Canadian contribution to global public health. *Can J Public Health* 2006; **97:** 42–44.

3   Brownstein JS, Freifeld CC. HealthMap: the development of automated real-time internet surveillance for epidemic intelligence. *Eurosurveillance* 2007; **12:** E071129 5.

4   Milinovich GJ, Williams GM, Clements AC, Hu W. Internet-based surveillance systems for monitoring emerging infectious diseases. *Lancet Infect Dis* 2014; **14:** 160–68.

5   WHO. Ebola response roadmap situation report—25 October 2014. http://apps.who.int/iris/bitstream/10665/137185/1/roadmapupdate25Oct14_eng.pdf?ua=1 (accessed Oct 26, 2014).