COACTION PUBLISHING

Web search query data to monitor dengue epidemics: a new model for dengue surveillance

Emily Chan[1], Vikram Sahai[2], Corrie Conrad[2] and John Brownstein[1]*

[1]Children's Hospital Boston, Boston, MA, USA; [2]Google, Inc., Mountain View, CA, USA

Objective

We aimed to assess whether web search queries are a viable data source for the early detection and monitoring of dengue epidemics.

Introduction

With an estimated 500 million people infected each year, dengue ranks as one of the most significant mosquito-borne viral human diseases and one of the most rapidly emerging vector-borne diseases (1). A variety of obstacles including bureaucracy and lack of resources have interfered with timely detection and reporting of dengue cases in many endemic countries (2). Surveillance efforts have turned to modern data sources, such as Internet search queries, which have been shown to be effective for monitoring influenza-like illnesses (3, 4). However, few have evaluated the utility of web search query data for other diseases, especially those of high morbidity and mortality or where a vaccine may not exist.

Methods

Bolivia, Brazil, India, Indonesia and Singapore were chosen for analysis based on data availability and adequate search volume. For each country, a univariate linear model was built by fitting a time series of the fraction of Google search query volume for specific dengue-related queries from that country against a 'gold standard' time series of dengue case counts for a time-frame within 2003–2010. The specific combination of queries used was chosen to maximize model fit. Spurious spikes in the data were also removed prior to model fitting. The final models, fit using a training subset of the data, were cross-validated against both the overall dataset and a holdout subset of the data. All search queries were fully anonymized. This methodology is similar to the approach used to develop Google flu trends (3).

Results

Dengue generated over a million Google search queries every month. Some queries showed that the user was looking for more information about the disease, while others were looking for symptoms or treatments. Model-fitted 'expected' epidemic curves matched official case counts 'observed' epidemic curves quite well for all 5 countries in most countries (Fig. 1), with validation correlations ranging from 0.82 to 0.99. Dengue queries were not as influenced by mass panic-induced searching.

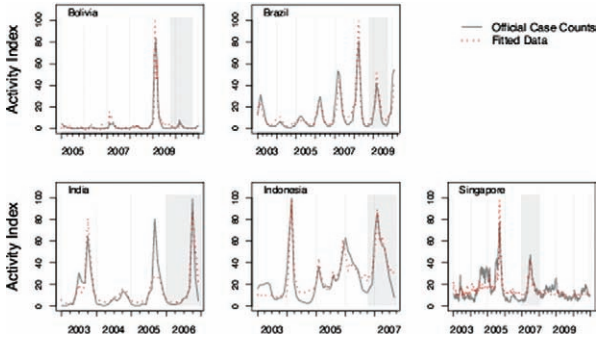

Fig. 1. A comparison of the model-fitted and official case counts dengue epidemic curves in each country.

Conclusions

Web search query data were found to be capable of tracking dengue activity in Bolivia, Brazil, India, Indonesia and Singapore. Whereas traditional dengue data from official sources are often not available until after some substantial delay, web search query data are available in near real-time and could serve as a useful low-cost complement to traditional surveillance. Even if peaks are no earlier, there is value in 'now-casting'—predicting the present where there are delays in gaining access to current official data (5). More broadly, these results also contribute to a growing pool of evidence demonstrating the capability of relatively novel sources such as web-based data to assist with public health goals. The product of this work is freely available at www.google.org/denguetrends.

Keywords

Dengue; epidemics; public health

References

1. Guzman MG, Halstead SB, Artsob H, Buchy P, Farrar J, et al. Dengue: a continuing global threat. Nat Rev Microbiol. 2010;8:S7–16.
2. Runge-Ranzinger S, Horstick O, Marx M, Kroeger A. What does dengue disease surveillance contribute to predicting and detecting outbreaks and describing trends? Trop Med Int Health. 2008;13:1022–41.
3. Ginsberg J, Mohebbi MH, Patel RS, Brammer L, Smolinski MS, et al. Detecting influenza epidemics using search engine query data. Nature. 2009;457:1012–4.
4. Polgreen PM, Chen Y, Pennock DM, Nelson FD. Using internet searches for influenza surveillance. Clin Infect Dis. 2008;47:1443–8.
5. Choi H, Varian H. Predicting the present with Google trends; 2009.

*John Brownstein
E-mail: john.brownstein@childrens.harvard.edu

Citation: Emerging Health Threats Journal 2011, **4**: 11135 - DOI: 10.3402/ehtj.v4i0.11135