COACTION PUBLISHING

Do geographic trends of social media indicate risk of secondary infectious disease outbreaks?

Rumi Chunara*, Marie Goetzke and John Brownstein

Harvard Medical School, Children's Hospital Boston, Boston, MA, USA

Objective

To evaluate the association between and develop a risk model relating geographic trends of social media and spread of an infectious disease outbreak.

Introduction

A devastating cholera outbreak began in Haiti in 2010. Sequencing of *Vibrio cholerae* isolates showed that the epidemic was likely the result of the introduction of cholera from a distant geographic source. The same strain of cholera was detected in other countries within 100 days. The unique instigation and geographic spread of this epidemic highlight the need for improvements in timely global outbreak surveillance. Novel information sources have been shown to provide early information about public health events and disease epidemiology. Particularly, volume of Internet metrics such as web searches or microblogs have been shown to be a good corollary for public health events (1). In this study, we evaluate geographic trends in online social media following an infectious disease outbreak to determine whether this may enable prediction of secondary outbreak locations.

Methods

We examined Twitter postings from the first 100 days of the Haitian cholera outbreak. Twitter is a microblogging service in which users can give information in 140 character length posts, 'Tweets'. We selected Tweets containing the word 'cholera' including those with the Twitter hashtag identifier ('#cholera'). Our search captured English, French and Spanish mentions of the word cholera. We define an outbreak as cholera incidence beyond an isolated case. Six countries in which cholera did or was suspected to have spread, and without endemic cholera, were examined: Canada, Dominican Republic, Mexico, Spain, USA and Venezuela. We first collected 'Twitter Updates' for each country, Tweets that came from users in a particular country, normalized by the number of Twitter users in the country. Second, we filtered Tweets in which the keyword cholera as well as a country's name was mentioned, 'Twitter Mentions'. Logistic models were constructed to analyze the relationship between volume of updates and mentions and the occurrence of a secondary cholera outbreak in the chosen countries. We evaluated our models through the Hosmer-Lemeshow (HL) test and also by cross-validation with data from Puerto Rico, in which there was concern of a potential outbreak.

Fig. 1. Distribution of Twitter updates in the 100 days following the Haiti cholera outbreak (green-red = low-high) and sample Tweets from Canada and the Dominican Republic.

Results

Global Tweets regarding a disease outbreak include concern from family or friends, local happenings and reiteration of news reports. Fig. 1 illustrates example Tweets and distribution of Twitter Updates in this study. The HL test yielded p-values of $\sim$ 1 and 0.185 (updates and mentions models). Large p-values indicate that the null hypothesis cannot be rejected and the model fits the expected distribution of values well, in this case, better for the updates model. Both models output a low probability (mentions: 0.04, updates: 6e-11) of an outbreak in Puerto Rico within 100 days, and there was no actual outbreak.

Conclusions

Global discussion of disease outbreaks may indicate where an outbreak will spread. This is the first study to examine how these discussions, via social media, can be used to understand and predict geographic spread of disease. Both the models demonstrated good fit to expected distributions through the HL test and correctly predicted no outbreak in Puerto Rico. We are working on incorporating data from more countries into the model, as well as other covariates such as environmental factors that would contribute to a country's tendency toward an outbreak. Although the global microblogging community is currently limited in demographics, penetrat\ion of consumer technology is increasing worldwide and could be a useful complementary tool for timely and cost-effective disease outbreak surveillance.

Keywords

Infectious disease; secondary outbreak; social media; informal surveillance

Reference

1. Chunara R, Andrews JR, Brownstein JS. (2011) Social and news media enable estimation of epidemiological patterns early in the 2010 Haitian cholera outbreak. Am J Trop Med Hyg. (in press).

*Rumi Chunara
E-mail: rumi@alum.mit.edu

Citation: Emerging Health Threats Journal 2011, **4**: 11110 - DOI: 10.3402/ehtj.v4i0.11110