OJPHI

ISDS 2015 Conference Abstracts

ISDS INTERNATIONAL SOCIETY FOR DISEASE SURVEILLANCE

# Social Media Analytics for Post-Disasters Disease Detection in the Philippines

**Lauren Charles-Smith*[3], Corinne Ringholz[1], Benjamin Brintz[3, 2] and Courtney D. Corley[3]**

[1]USAID, Washington, DC, USA; [2]Oregon State University, Corvallis, OR, USA; [3]Pacific Northwest National Laboratory, Richland, WA, USA

## Objective

To determine the potential of Twitter data as an early warning of a likely communicable disease outbreak following a natural disaster, and if successful, develop an open-source algorithm for use by interested parties.

## Introduction

Previous research identifies social media as an informal source of near-real time health data that may add value to disease surveillance systems by providing broader access to health data across hard-to-reach populations. This indirect health monitoring may improve public health professionals' ability to detect disease outbreaks faster than traditional methods and to enhance outbreak response. The Philippines consists of over 7,000 islands and is prone to meteorological (storms), hydrological (floods), and geophysical disasters (earthquakes and volcanoes). In these situations, evacuation centers are used for safety and medical attention and often house up to 50K people each for 2 or more months, sometimes with unclean water sources and improper sanitation. Consequently, these conditions are a perfect venue for communicable disease transmission and have been proposed to cause disease outbreaks weeks after the original disaster occurred. Coined the social media capital of the world[1], the Philippines provides a perfect opportunity to evaluate the potential of social media use in disease surveillance.

## Methods

The social media analyzed consists of 50 million geotagged tweets from the Philippines between 2012 and 2013. Monthly disease outbreak case counts by location were collected from the Filipino Department of Health. Disaster data was identified through the National Operational Assessment of Hazards and the International disaster database website, EM-Dat. All data were split into 17 national regions to improve spatial resolution and decrease social variation within populations. Outbreaks of interest were identified if they occurred 1-2 months following a natural disaster. Topic modeling and theme identification methods were used to explore and understand Filipino Twitter use and language. To identify tweets of interest, lexicons were developed in English, Tagalog, Taglish, and other native dialects. The final disease lexicon, verified by visual confirmation, was used to filter the tweets and create histograms of tweet counts per day per region. This data was modeled by time series analyses to identify change points indicative of disease outbreaks (R BreakoutDetection). The data was analyzed on multiple spatial scales and compared to known disease outbreak counts and natural disasters with 4 potential spatiotemporal correlations between disease and disaster identified. Autoregressive integrated moving average model was used to forecast spikes and enhance outbreak detection (Fig. 1). This tweet model identified 1 potential correlation between tweet, disaster, and outbreak in 2012 and 4 in 2013, potentially illustrating the increase of power in social media as twitter use increases over time. Monthly outbreaks were regressed against various predictors including tweets (past and current), disasters (type, time since last occurrence), and region-specific characteristics (population density).

## Conclusions

Current models using monthly disease outbreak data lack significant correlation, which is most likely due to the loss of information when aggregating data to a monthly scale. In effort to increase outbreak detection, models will be developed using newly acquired continuous weekly case counts by region. In addition, disease-specific outbreaks will be regressed against disease-specific filtered tweets and disaster information to determine the best combination of predictors by region. The final goal is to create a model that integrates historical disaster data with disease-related Twitter counts to be used as a disease forecast system for streaming Twitter data. This effort, funded by USAID, will be transitioned to regional universities and the Government of the Philippines.

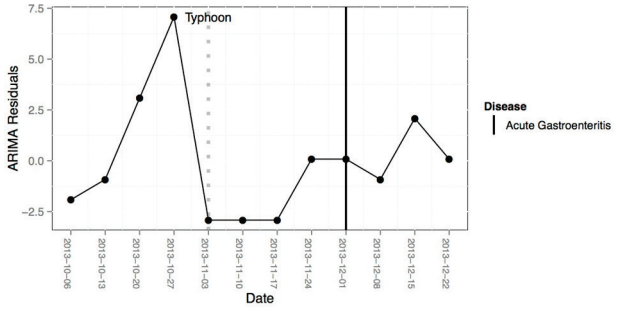


Figure 1. Histogram of disease tweets and acute gastroenteritis outbreak post Typhoon Yolanda in the Philippines Eastern Visayas region.

## Keywords

Health surveillance; disasters; social media; Biosurveillance

## Acknowledgments

This effort was funded by the USAID.

## References

[1] Boaz JL, Ybañez M, De Leon MM, et al. Understanding the Behavior of Filipino Twitter Users during Disaster. GSTF J Comput 2013 DOI: 10.7603/s40601-013-0007-z

***Lauren Charles-Smith**
E-mail: lauren.charles-smith@pnnl.gov