

Embargoed until 9 a.m. (EDT) August 12, 2015

STATISTICIANS USING SOCIAL MEDIA TO TRACK FOODBORNE ILLNESS OUTBREAKS AND IMPROVE DISASTER RESPONSE

SEATTLE, WA, AUGUST 12, 2015 – The growing popularity and use of social media around the world is presenting new opportunities for statisticians to glean insightful information from the infinite stream of posts, tweets and other online communications that will help improve public safety.

Two such examples—one that enhances systems to track foodborne illness outbreaks and another designed to improve disaster-response activities—were presented this week at the [2015 Joint Statistical Meetings](#) (JSM 2015) in Seattle.

Tracking Foodborne Illness Outbreaks

In a presentation titled “[Digital Surveillance of Foodborne Illnesses and Outbreaks](#)” presented yesterday, biostatistician Elaine Nsoesie unveiled a method for tracking foodborne illness and disease outbreaks using social media sites such as Twitter and business review sites such as Yelp to supplement traditional surveillance systems. Nsoesie is a research fellow in pediatrics at Boston Children’s Hospital.

The study’s purpose was to assess whether crowdsourcing via online reviews of restaurants and other foodservice institutions can be used as a surveillance tool to augment the efforts of local public health departments. These traditional surveillance systems capture only a fraction of the estimated 48 million foodborne illness cases in the country each year, primarily because few affected individuals seek medical care or report their condition to the appropriate authorities.

Nsoesie and collaborators tested their nontraditional approach to track these outbreaks. The results showed foods—for example, poultry, leafy lettuce and mollusks—implicated in foodborne illness reports on Yelp were similar to those reported in outbreak reports issued by the U.S. Centers for Disease Control and Prevention.

“Online reviews of foodservice businesses offer a unique resource for disease surveillance. Similar to notification or complaint systems, reports of foodborne illness on review sites could serve as early indicators of foodborne disease outbreaks and spur investigation by local health authorities. Information gleaned from such novel data streams could aid traditional surveillance systems in near real-time monitoring of foodborne related illnesses,” said Nsoesie.

The lack of near real-time reports of foodborne outbreaks reinforces the need for alternative data sources to supplement traditional approaches to foodborne disease surveillance, explained Nsoesie. She added Yelp.com data can be combined with additional data from other social media sites and crowdsourced websites to further improve coverage of foodborne disease reports.

Enhancing Disaster Response by Analyzing Social Media

As part of a team of statisticians from [Statistics without Borders](#) (SWB)—an outreach group of the American Statistical Association—Michiko Wolcott and several colleagues evaluated social media traffic posted during and the days following Typhoon Haiyan striking the Philippines in November 2013 to develop a set of social media analytics best practices for emergency response managers.

The project was conducted in coordination with [Humanity Road](#), a volunteer-based charity that delivers disaster preparedness and response information to the global mobile public before, during and after a disaster. The collaboration led to the development of an informational resource for emergency management professionals titled “[A Guide to Social Media Emergency Management Analytics](#).” SWB and Humanity Road are both members of the Digital Humanitarian Network, consisting of volunteer and nonprofit organizations that support leveraging of digital technology in humanitarian response situations.

Wolcott today presented a summary of SWB’s recent work with DHN network organizations, as well as the findings and key recommendations in the guidebook during an invited presentation titled “[Worldwide Statistics without Borders Projects: SWB Helping Organizations Make Better Decisions](#).”

The project’s overall objective was to analyze the tweets to identify best practices for data handling, identify analysis approaches for emergency response and recommend data management approaches. Important considerations and challenges were identified regarding the use and analysis of Twitter-based data sets for disaster response, noted Wolcott.

“Social media can play a critical role in the dissemination of the information, as well as collection of relevant data during natural disasters. The idea of leveraging social media data such as Twitter is intuitively attractive, given their natural ties to mobile devices with obvious disaster response implications,” explained Wolcott.

The guidebook notes there are a number of key considerations to ensure the analysis of social media during a natural disaster is designed to meet the objective. The opportunity for data analysis must be properly and promptly identified, and the disaster response resources and analytical resources must work together to determine how to best house, extract and analyze the data.

Among the recommendations for analysis of social media included in the guidebook are the following:

1. **Relevance**—Filtering criteria such as country, keywords, hashtags, geolocation, language, type of posts (e.g., organic vs. retweets) and type of poster (e.g., individuals, relief organizations, news organizations, celebrities, etc.) must be carefully considered based on the analysis objective.

2. **Geolocatability**—In many cases, the basis of all insight from social media posts is the geolocatability of the tweets. Only a portion of the relevant tweets were geolocated, and, of those, 37% come from the Philippines, 25% from the United States, followed by tweets from Great Britain, Canada and Vietnam. While these results show a global interest and awareness in the event, factors such as the proportion of geolocated tweets and the method of geolocation plays an important part in the decisions regarding geolocation. Capturing the specific motivation for the tweets will depend on the analysis objective, with important implications in the design of data collection to which emergency management professionals and analysts must be sensitive.
3. **Language**—The particularities of Twitter make language identification challenging—the length of messages, heavy use of hashtags and abbreviations and variations in users’ communication styles require considerations beyond straightforward use of standard language identification tools. Furthermore, analyzing social media data in countries such as the Philippines presents additional challenges because the country’s residents use several languages and issues such as variations in transliteration can add to the challenges. Emergency management professionals and analysts must be prepared to address these issues.
4. **Device vs. impact of the disaster on infrastructure**—The penetration and usage for a particular device/platform varies by region and country, which must be taken into consideration. Furthermore, emergency management professionals must be sensitive to interruptions in electricity and communications infrastructure because these may affect the data.

The guidebook also offers a list of questions that will help emergency management professionals start a dialogue about social media emergency management analysis. Broad areas covered in the questions are the handling and storage of data; creating a baseline and identifying the type of content and trends; and planning the reporting time window, location and language.

In recognition of its work on this project, SWB was honored with Humanity Road’s 2014 Da Vinci Award, presented to a patron or contributor who supports the organization’s programs.

JSM 2015 is being held August 8–13 at the Washington Convention Center. More than 6,000 statisticians—representing academia, business and industry, as well as national, state and local governments—from numerous countries are attending North America’s largest statistical science gathering.

About JSM 2015

JSM, which has been held annually since 1974, is being conducted jointly this year by the [American Statistical Association](#), [International Biometric Society \(ENAR and WNAR\)](#), [Institute of Mathematical Statistics](#), [Statistical Society of Canada](#), [International Chinese Statistical Association](#), [International Indian Statistical Association](#), [Korean International Statistical Society](#), [International Society for Bayesian Analysis](#), [Royal Statistical Society](#), and [International Statistical Institute](#). JSM activities include oral presentations, panel sessions, poster presentations, continuing professional development courses, an exhibit hall, a career service, society and section business meetings, committee meetings, social activities and networking opportunities. [Click here for more information about JSM 2015.](#)

About the American Statistical Association

The ASA is the world's largest community of statisticians and the second-oldest continuously operating professional society in the United States. Its members serve in industry, government and academia in more than 90 countries, advancing research and promoting sound statistical practice to inform public policy and improve human welfare. For additional information, please visit the ASA website at www.amstat.org.

###

For more information or to request an interview with Elaine Nsoesie or Michiko Wolcott:

Jeffrey A. Myers

Office: (703) 684-1221, Ext. 1865

Mobile: (540) 623-7777

Email: jeffrey@amstat.org