

# Evaluating Twitter for Foodborne Illness Outbreak Detection in New York City

Katelynn Devinney<sup>\*1</sup>, Adile Bekbay<sup>1</sup>, Thomas Effland<sup>2</sup>, Luis Gravano<sup>2</sup>, David Howell<sup>1</sup>, Daniel Hsu<sup>2</sup>, Daniel O'Hallorhan<sup>1</sup>, Vasudha Reddy<sup>1</sup>, Faina Stavinsky<sup>1</sup>, HaeNa Waechter<sup>1</sup> and Bruce Gutelius<sup>1</sup>

<sup>1</sup>Bureau of Communicable Disease, NYC Department of Health and Mental Hygiene, Queens, NY, USA; <sup>2</sup>Columbia University, New York, NY, USA

## Objective

To incorporate data from Twitter into the New York City Department of Health and Mental Hygiene foodborne illness surveillance system and evaluate its utility and impact on foodborne illness complaint and outbreak detection.

## Introduction

An estimated one in six Americans experience illness from the consumption of contaminated food (foodborne illness) annually; most are neither diagnosed nor reported to health departments<sup>1</sup>. Eating food prepared outside of the home is an established risk factor for foodborne illness<sup>2</sup>. New York City (NYC) has approximately 24,000 restaurants and >8.5 million residents, of whom 78% report eating food prepared outside of the home at least once per week<sup>3</sup>. Residents and visitors can report incidents of restaurant-associated foodborne illness to a citywide non-emergency information service, 311. In 2012, the NYC Department of Health and Mental Hygiene (DOHMH) began collaborating with Columbia University to improve the detection of restaurant-associated foodborne illness complaints using a machine learning algorithm and a daily feed of Yelp reviews to identify reports of foodborne illness<sup>4</sup>. Annually, DOHMH manages over 4,000 restaurant-associated foodborne illness reports received via 311 and identified on Yelp which lead to the detection of about 30 outbreaks associated with a restaurant in NYC. Given the small number of foodborne illness outbreaks identified, it is probable that many restaurant-associated foodborne illness incidents remain unreported. DOHMH sought to incorporate and evaluate an additional data source, Twitter, to enhance foodborne illness complaint and outbreak detection efforts in NYC.

## Methods

DOHMH epidemiologists continue to collaborate with computer scientists at Columbia University who developed a text mining algorithm that identifies tweets indicating foodborne illness. Twitter data are received via a targeted application program interface query that searches for foodborne illness key words and uses metadata to select for tweets with a possible NYC location. Each tweet is assigned a sick score between 0–1; those meeting a threshold value of 0.5 are manually reviewed by an epidemiologist, and a survey link is tweeted to users who have tweeted about foodborne illness, requesting more information regarding the date and time of the foodborne illness event, restaurant details, and user contact information. Survey data are used to validate complaints and are incorporated in a daily analysis using all sources of complaint data to identify restaurants with multiple foodborne illness complaints within a 30-day period. This system was launched on November 29, 2016.

## Results

During November 29, 2016–September 27, 2017, 12,015 tweets qualified for review (39/day on average); 2,288 (19.0%) indicated foodborne illness in NYC, and 1,778 (14.8%) were tweeted a survey link (510 foodborne illness tweets were either deleted by the Twitter

user or were tweets from a user who was already sent a survey for the same foodborne illness incident). The survey tweets resulted in 92 likes, 12 retweets, 65 replies, 232 profile views and 348 survey link clicks. Of the 1,778 surveys sent, 27 were completed (response rate 1.5%), of which 20 (74.7%) confirmed foodborne illness associated with a NYC restaurant; none had been reported via 311/Yelp. Of those, 11 (55%) provided a phone number, of which 10 (90.9%) completed phone interviews. The completed surveys contributed to the identification of two restaurants with multiple foodborne illness complaints within a 30-day period.

## Conclusions

The utility of Twitter for foodborne illness outbreak detection continues to be evaluated. While the survey response rate has been low, the identification of new complaints not otherwise reported to 311 and Yelp suggests this will be a useful tool. Future plans include using feedback data collected by DOHMH epidemiologist review to increase the sensitivity and specificity of the text mining algorithm and improve the location detection for Twitter users. In addition, we plan to implement enhancements to the survey and create a web page to promote survey responses. Furthermore, we intend to share this system with other health departments so that they might incorporate Twitter in their outbreak detection and public health surveillance activities.

## Keywords

Social media; outbreak detection; text mining

## References

1. Scallan E, Griffin PM, Angulo FJ, Tauxe RV, Hoekstra RM. Foodborne illness acquired in the United States--unspecified agents. *Emerg Infect Dis.* 2011 Jan;17(1):16-22.
2. Jones TF, Angulo FJ. Eating in restaurants: a risk factor for foodborne disease? *Clin Infect Dis.* 2006 Nov 15;43(10):1324-8.
3. New York City Health and Nutrition Examination Survey, 2013-2014 [Internet]. New York: New York City Department of Health and Mental Hygiene and The City University of New York; 2017 [cited 2017 Aug 28]. Available from: <http://nychanes.org/data/>
4. Harrison C, Jorder M, Stern H, Stavinsky F, Reddy V, Hanson H, Waechter H, Lowe L, Gravano L, Balter S; Centers for Disease Control and Prevention (CDC). Using online reviews by restaurant patrons to identify unreported cases of foodborne illness - New York City, 2012-2013. *MMWR Morb Mortal Wkly Rep.* 2014 May 23;63(20):441-5.

\*Katelynn Devinney

E-mail: [kdevinney@health.nyc.gov](mailto:kdevinney@health.nyc.gov)

