Improving Foodborne Complaint and Outbreak Detection Using Social Media, New York City

Arizona Infectious Disease Training, July 19, 2017
Foodborne Illness

• Illness caused by the consumption of contaminated food
• Many different pathogens
• Vastly underdiagnosed and not reported to health departments
• Estimated 48 million episodes of foodborne illness in United States each year\(^1\)
• Eating food prepared outside of the home increases risk\(^2\)


Foodborne Illness Complaints in New York City (NYC)

- 24,000 restaurants and 15,000 food retailers
- Over 8.5 million residents; 78% report going out to eat ≥ once per week³
- NYC’s non-emergency information service - 311
  - Allows residents to submit various types of complaints, including food poisoning
  - Submitted by phone or online
- NYC Department of Health and Mental Hygiene (DOHMH) receives ~3,500 restaurant associated complaints each year
- Approx. 30 restaurant related outbreaks each year

Unreported Foodborne Illness Complaints

• 2011: Outbreak investigation identified multiple complaints on the restaurant review site, Yelp.com that were not submitted via 311

• Nine month pilot project launched in 2012
  • Found 468 reviews consistent with foodborne illness; identified 3 previously unreported outbreaks
  • Many reviewers unaware of 311

• Collaborated with Yelp and Columbia University to obtain daily feed of Yelp reviews and develop a machine learning program to classify and identify reviews pertaining to foodborne illness

Yelp Integration

• Nightly process to pull data from Yelp via Amazon Web Services; Google geocoding used to select only for NYC restaurants.

• Program developed by Columbia assigns a “sick score” indicating the likelihood that a review pertains to foodborne illness; ranges from 0-1.0 and is based on if the review indicates:
  • Food poisoning or symptoms of foodborne illness (vomit, diarrhea, etc.)
  • Multiple people became sick
  • Incubation period
Yelp Integration

• All reviews enter the Foodborne Illness Tracker application; those with sick score $\geq 0.5$ manually classified by DOHMH staff

• Messages sent to those identified as complaints, requesting they contact DOHMH for interview

• Feedback data sent to Columbia
## Yelp Example

**Pros:** Good food and friendly service! Takes reservations online! Convenient location! **Cons:** Ate an early dinner here before going to see Hamilton. Food poisoning hit during intermission because I could not leave during intermission because I could not stop being various forms of sick and shaking. Really bummed I missed the second half of Hamilton! Maybe only go if you don't have $300 theater tickets that night. Heed my mistake.

| Is this a case of foodborne disease: | Yes |
| Did it occur within the last month: | Unknown |
| Did two or more people become ill: | No |
| Are there severe symptoms: | Yes |
| Action item: | Urgent |
| What type of message was sent: | Interview message |

**Rating:** 0.886
## Yelp Response

<table>
<thead>
<tr>
<th></th>
<th>Number</th>
<th>Percent</th>
</tr>
</thead>
<tbody>
<tr>
<td>Reviews</td>
<td>17314</td>
<td>100</td>
</tr>
<tr>
<td>Foodborne</td>
<td>8770</td>
<td>50.7</td>
</tr>
<tr>
<td>Messaged</td>
<td>4401</td>
<td>59.3</td>
</tr>
<tr>
<td>Responded</td>
<td>1084</td>
<td>24.6</td>
</tr>
<tr>
<td>Interviewed</td>
<td>527</td>
<td>12.0</td>
</tr>
</tbody>
</table>

Data for Yelp reviews from July 1, 2012- July 12, 2017

- Since 2012, 12 outbreaks identified using Yelp
Unreported Foodborne Illness Complaints

- Even with the inclusion of Yelp, still likely not receiving all complaints of restaurant associated foodborne illness in NYC
- Other jurisdictions have reported success in identifying foodborne illness complaints using Twitter
- We sought to:
  - Identify tweets indicating foodborne illness
  - Validate tweets using an online survey
  - Integrate completed surveys into foodborne illness complaint system
Twitter Integration

- Use publicly available Twitter application program interface (API)
- Data received every two hours via a targeted API query that searches for keywords pertaining to foodborne illness
- Tweets assigned a sick score (0-1.0); solely based on indication of illness, food poisoning, symptoms, etc.
- Location obtained via metadata; based on user account registration
Twitter Integration

• Tweets with sick score $\geq 0.5$ manually reviewed by DOHMH staff
• Survey link tweeted back to users complaining of foodborne illness
• Completed surveys qualify as complaints
• Feedback data sent to Columbia
**Twitter Example**

<table>
<thead>
<tr>
<th>Tweet Text</th>
<th>Sick Score</th>
<th>Location</th>
<th>Comment</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Got a bad case of food poisoning on Friday. How long before I’m not afraid to eat food again?</strong></td>
<td>0.612</td>
<td>Bronx, NY</td>
<td></td>
</tr>
<tr>
<td><strong>I felt sick to my stomach hearing about Kim’s story. She had 2 young children. Rich or not. Totally devastating. Folks were callus. Smdh.</strong></td>
<td>0.974</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
We found your tweet about possible food poisoning and would like to ask you some questions.

If you think you have food poisoning in New York City, please complete this form. The info will be sent to the NYC Dept. of Health and Mental Hygiene so they can take any necessary action.

What happened?
Please be as descriptive as possible, letting us know what food items you consumed, what time of day, what was your first symptom, and for how long you were sick.

* What restaurant was it?
  e.g. Restaurant Name

* Where was the restaurant located?
  e.g. 42-09 28th St, Queens, NY 11101
  Please provide address, town, or cross streets

* What happened?
  e.g. I went to [Restaurant Name] on [Date] at [Time] with 4 friends. I ate the chicken salad sandwich and tomato soup. The next morning I experienced severe diarrhea and vomiting and felt sick for about 24 hours.

What day did you order food from the restaurant?

How can the City reach you?

First name: Jane

Last name: Doe

Email: janedoe@example.com

Phone: 212-555-5555

Submit your report
## Twitter Survey Response

<table>
<thead>
<tr>
<th></th>
<th>Number</th>
<th>Percent</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tweets reviewed</td>
<td>9754</td>
<td>100</td>
</tr>
<tr>
<td>Foodborne and NYC</td>
<td>2444</td>
<td>25.1</td>
</tr>
<tr>
<td>Survey sent</td>
<td>2046</td>
<td>21.0</td>
</tr>
<tr>
<td>Survey completed</td>
<td>23</td>
<td>1.1</td>
</tr>
<tr>
<td>Interview completed</td>
<td>9</td>
<td>39</td>
</tr>
</tbody>
</table>

Data for tweets identified 11/29/16-7/12/2017

- Of 23 completed surveys:
  - 4 outside NYC; one for an item purchased at grocery store
  - 6 did not provide complete contact information (phone number and email)
  - Of 12 in NYC with contact information, 9 (75%) completed interviews
# Twitter Public Response

<table>
<thead>
<tr>
<th></th>
<th>Number</th>
<th>Percent</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tweets sent</td>
<td>2046</td>
<td>100</td>
</tr>
<tr>
<td>Likes</td>
<td>75</td>
<td>3.7</td>
</tr>
<tr>
<td>Retweets</td>
<td>11</td>
<td>0.5</td>
</tr>
<tr>
<td>Replies</td>
<td>55</td>
<td>2.7</td>
</tr>
<tr>
<td>Survey link clicks</td>
<td>333</td>
<td>16.3</td>
</tr>
<tr>
<td>Profile views</td>
<td>210</td>
<td>10.3</td>
</tr>
<tr>
<td>Detail expansions</td>
<td>940</td>
<td>45.9</td>
</tr>
</tbody>
</table>

Data for tweets identified 11/29/16-7/12/2017
Engagement Promotion

- Survey response rate very low (1.1%)
- Implemented changes to promote engagement and survey completion
  - Tweet more frequently to increase Twitter presence
  - Edit response message
  - Add infographic
• Please complete this survey to report illness to the NYC Health Dept: https://t.co/L60xqSb7nE

• Sorry that you're sick! Please complete this survey to report illness to the NYC Health Dept: https://t.co/TRB0I54Llw

• Hope you're feeling better! Help the NYC Health Dept. prevent this from happening again: https://t.co/CfqbMkdXiY

• Hope you are feeling better! Help the NYC Health Dept. prevent #foodpoisoning from happening again: https://t.co/Lzb0NC3ngd
The NYC Dept. of Health found your tweet about possible food poisoning and would like to ask you a few questions.

If you think you have food poisoning in New York City, please complete this form. The info will be sent to the NYC Dept. of Health and Mental Hygiene so we can take any necessary action to prevent others from becoming sick.
Daily Report

• SAS is used to generate and email a daily report; aggregates data from 311, Yelp, and Twitter surveys
• Complaints are matched based on restaurant name and location
• Restaurants with multiple complaints within a 30 day period are flagged for investigation
Implementation Requirements

• Supported by grants from Alfred P. Sloan Foundation and National Science Foundation

• Collaboration with Columbia University Department of Computer Science

• Additional Staff
  • Programmer/developer
  • Epidemiologist
  • Student intern
  • Public health inspector

• Coordination with IT, ongoing maintenance
Challenges

• Technical difficulties
  • Firewall issues, access to data blocked by IT security
  • Public facing server with increased IT security required for survey

• Twitter
  • Geolocation data no longer provided in public API
  • Anti-spamming measures
  • Difficult to contact, less willing to collaborate

• Noise/background

• Engagement
Next Steps

• Implement improved classifier for Yelp reviews
• Use Twitter feedback data to improve sensitivity and specificity of classifier
• Incorporate additional data sources
• Evaluate methods to increase response rates
Questions?

Katelynn Devinney, MPH
Bureau of Communicable Disease
NYC Department of Health and Mental Hygiene
kdevinney@health.nyc.gov