Investigating foodborne outbreaks

Sung Im
Bioinformatician
IHRC, Inc.

Center’s for Disease Control and Prevention, Atlanta, GA
National Centers for Emerging Zoonotic Infectious Diseases
Division of Foodborne and Waterborne Infections Diseases
Enteric Diseases Laboratory Branch
Bioinformatics and Metagenomics Subtyping Team
Impact of foodborne diseases

Each year in the U.S.
- 48 million illness
- 1 in 6 Americans
- 128,000 Hospitalizations
- 3000 Deaths
- $15 Billion

Each year Worldwide:
- 550 million illness
- 1 in 10 persons
- 230,000 Deaths
What we know

Transitioning the PulseNet USA network to whole genome sequencing: the final push

• WGS provides more resolution to identify outbreak clusters
• WGS data gives outbreak subtyping and genotyping information through the same [single] workflow
• WGS is just the first step for PulseNet, next stop is culture independent subtyping methods

Genomic epidemiology (algorithms and software)

• WGS provides high resolution
• We have many tools for differing levels of resolution
• We can and have used it on outbreak investigations
Disclaimers

• The findings and conclusions in this report are those of the authors and do not necessarily represent the official position of the Centers for Disease Control and Prevention

• I am not an epidemiologist
What is epidemiology?

Epidemiology is the study of the distribution and determinants of health-related states or events in specified populations and the application of this study to the control of health problems. [1]

- Data driven, quantitative discipline
- Systematic and unbiased approach to data collection, analysis, interpretation
- Draws on methods from other fields including biostatistics, [bio]informatics, and biologic, economic, social, behavior sciences
To tell a good story...

• What – diagnosis or health event
• Who – person(s) affected
• Where – place
• When – time
• Why/How – causes, risk factors, and modes of transmission
Distribution

**Frequency** – refers to number of health events, i.e. number of cases of *E. coli* infections in a population

**Pattern** – refers to the occurrence of health-related events by time, place, and person

- Time patterns may be annual, seasonal, weekly, daily, hourly, etc.
- Place patterns include geographic variations, urban/rural differences, etc.
- Personal characteristics include demographic factors, such as age, sex, marital status, an social economic class
Determinants

Determinant – refers to any factor, whether event, characteristic, or other definable entity, that brings about a change in a health condition (or other defined characteristic)

Epidemiologist use analytic epidemiology studies to search for determinants

- Explains the “why” and “how” component of the event
- Do groups with different rates of disease differ in their demographic characteristics, genetic or immunological make-up, behaviors, environmental exposure, or other potential risk factors
Application

- Epidemiology is not *just* the study of health in a population
- It also involves applying the knowledge gained by the studies to community-based practice
  - Just like a doctor combines medical knowledge with experience, clinical judgement, and understanding to diagnose and prescribe treatment
- Ideally, the findings provide sufficient evidence to direct prompt and effective public health control and prevention measures
Jon Snow
Jon Snow
John Snow - the father of modern epidemiology
1854 London Cholera outbreak

From John Snow’s On the mode of communication of cholera (2nd edition, 1855) © London School of Hygiene & Tropical Medicine Library & Archives
1854 London Cholera outbreak

From John Snow's *On the mode of communication of cholera* (2nd edition, 1855) © London School of Hygiene & Tropical Medicine Library & Archives
1854 London Cholera outbreak

From John Snow's On the mode of communication of cholera (2nd edition, 1855) © London School of Hygiene & Tropical Medicine Library & Archives
In summary

• Epidemiology is the study of the **distribution** and **determinants** of health-related states or events in specified populations and the **application** of this study to the control of health problems. [1]

• Beer > Water
Steps in a foodborne outbreak investigation

1. DETECT
   Detect a possible outbreak through public health surveillance.

2. FIND
   Find more cases in the outbreak.

3. GENERATE
   Generate hypotheses through interviews with sick people.

4. TEST
   Test hypotheses to find a likely source. If no source is found and cases continue, return to step 3.

5. SOLVE
   Solve source of the outbreak and ultimate point of contamination.

6. CONTROL
   Control outbreak through recalls, facility improvements, and industry collaboration.

7. DECIDE
   Decide an outbreak is over and the public is no longer at risk. If cases go up again, continue or restart the investigation.

Foodborne outbreak investigations are dynamic. In reality, some steps may happen at the same time.

Adapted from: Accessible version: https://www.cdc.gov/foodsafety/outbreaks/investigating-outbreaks/investigations/decision.html
Step 1: Detect an emerging outbreak

Detection is the first step in investigating a multistate foodborne outbreak

- An outbreak with multiple sick people can be missed if they are spread out over a wide area
- Outbreaks are detected by using public health surveillance methods (PulseNet), and formal/informal reports of illnesses

PulseNet connects the dots to detect foodborne outbreaks and prevent over 270,000 illnesses from Salmonella, E. coli and Listeria every year.
How are emerging outbreaks uncovered?

Local level
Exposure information from individual case reports are regularly reviewed to look for common factors.

State level
Outbreaks may be detected by regular, timely analysis of surveillance data that reveals an increase in reported cases or an unusual clustering of cases by time and place.

National level
Analyze surveillance data – reports of cases of communicable disease that are routinely sent by laboratories and health care providers to health departments.
# Guide to an confirming etiology in foodborne outbreak

<table>
<thead>
<tr>
<th>Etiologic Agent</th>
<th>Incubation Period</th>
<th>Search Window</th>
<th>Clinical Syndrome</th>
<th>Confirmation</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>Listeria monocytogenes</em> – Diarrheal disease</td>
<td>Unknown</td>
<td>120 days</td>
<td>Diarrhea, abdominal cramps, fever</td>
<td>Isolation of organism of same serotype from stool of <strong>two or more</strong> ill persons exposed to food that is epidemiologically implicated or from which organism of same serotype has been isolated</td>
</tr>
<tr>
<td><em>Escherichia coli</em> – Enterohemorrhagic (<em>E. coli</em> O157:H7 and others)</td>
<td>1-10 days; usually 3-4 days</td>
<td>60 days</td>
<td>Diarrhea (often bloody), abdominal cramps (often severe), little or no fever</td>
<td>Isolation of <em>E. coli</em> O157:H7 or other Shiga-like toxin-producing <em>E. coli</em> from clinical specimen from <strong>two or more</strong> ill persons</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>OR</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Isolation of <em>E. coli</em> O157:H7 or other Shiga-like toxin-producing <em>E. coli</em> from epidemiologically implicated food</td>
</tr>
<tr>
<td>Nontyphoidal <em>Salmonella</em></td>
<td>6hrs-10 days; usually 7-14 days</td>
<td>60 days</td>
<td>Diarrhea, often with fever and abdominal cramps</td>
<td>Isolation of organism of same serotype from clinical specimens from <strong>two or more</strong> ill persons</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>OR</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Isolation of organism from epidemiologically implicated food</td>
</tr>
</tbody>
</table>

[https://www.cdc.gov/foodafety](https://www.cdc.gov/foodsafety)
Step 2: Define and find cases

Finding sick people is important to understand the size, timing, severity, and possible sources of an outbreak

- A case definition is developed to define who will be included as part of an outbreak
- Epidemiologists use the case definition to search for illnesses related to the outbreak
- Illnesses are plotted on an epidemic curve (epi curve) to track when illnesses occur over time
Case definition

A standard set of criteria for deciding whether an individual should be classified as having the disease of interest

- Includes clinical criteria and restrictions on time, place, and person
- Standardization ensure comparability
Line list

<table>
<thead>
<tr>
<th>CaseID</th>
<th>Case Initials</th>
<th>Age</th>
<th>Sex</th>
<th>Onset Date</th>
<th>Current Status</th>
<th>Location</th>
<th>Case Category</th>
<th>Epi Links</th>
<th>Underlying Conditions</th>
</tr>
</thead>
</table>

*Page 4 contains a description of the column headings

**Line List Template**

- **CaseID:** Unique identifier assigned to each case patient for this investigation
- **Case Initials:** Case-patient initials
- **Age:** Age in years
- **Sex:** Male, Female or Unknown
- **Onset Date:** Date of symptom onset, ended by
- **Current Status:** Outpatient, Inpatient, Inpatient ICU, Discharged, Died
- **Location:** Hospital, City, County
- **Case Category:** Confirmed, probable, suspect
- **Epi Links:** Knowns exposures, affiliations or connections to other cases
- **Underlying Conditions:** Significant immunodeficiencies, medications or other conditions that may alter the patient’s susceptibility or course
- **Radiology:** Was a chest x-ray performed? If so, what were the results?
- **Specimen collected:** Examples: Nasopharyngeal (NP) swab, nasopharyngeal wash/irrigation, endotracheal (ET) swab, sputum, tracheal aspirates, bronchoalveolar lavage (BAL), pleural fluid, lung tissue, blood, ascites, urine
- **Testing performed:** Examples: Culture, antigen detection, antibody serology, polymerase chain reaction (PCR), immunohistochemistry (IHC)
- **Results:** Findings of laboratory testing to date

*Indicate specimen type for each test or result documented
Step 3: Generate hypothesis about likely sources

Hypothesis generation is an ongoing process

- Possible explanations of an outbreak are continually changed or disproved as more information is gathered
- Interviews, questionnaires, and home visits are helpful in narrowing down how and where case patients got sick
Hypothesis generating questionnaire

What is the Hypothesis generating questionnaire?
The National Hypothesis Generating Questionnaire is a set of questions used by public health officials to interview ill people in the early stages of multistate foodborne or enteric (gastrointestinal) disease outbreak investigation.

Why is it used?
The questionnaire collects a standard set of information about food or other exposures for all outbreak cases identified during a multistate investigations. By collecting the same information across many different areas, data analysis is more efficient. Also, the time it takes to pinpoint the source of an outbreak can be reduced.

What happens to the data?
De-identified data collected by public health officials are usually combined and analyzed by CDC and the results are shared with federal, state, and local public health partners, including the U.S. Food and Drug Administration and the U.S. Department of Agriculture.
Descriptive epidemiology

Characterization of health events by **time**, **place**, and **person**

**Pattern** – refers to the occurrence of health-related events by time, place, and person

- Time patterns may be annual, seasonal, weekly, daily, hourly, etc.
- Place patterns include geographic variations, urban/rural differences, etc.
- Personal characteristics include demographic factors, such as age, sex, marital status, and social economic class
Descriptive epidemiology: time

Time patterns

- The occurrence of disease changes over time
- Studying time graphs (histograms) that note the period of exposure can lead to insights into what may have caused the illness
- The shape and other features of an epi curve can suggest hypotheses about the time and source of exposure, the mode of transmission, and the causative agent
Descriptive epidemiology: place

Place patterns

- Provides insight into the geographic extent of the problem
- Can refer to place of residence, place of diagnosis or report, birthplace, site of employment, school district, hospital unit, or recent travel destinations
Descriptive epidemiology: person

Personal characteristic patterns

Age
- Almost every health-related event varies with age
- Says a lot about susceptibility, opportunity for exposure, latency in incubation period, physiological response

Sex
- Males have higher rates of illness and death than do females for many diseases
- Sex differences affect susceptibility and physiological responses
Step 4: Test hypothesis

A hypothesis is tested to determine if the outbreak source has been correctly identified

- Analytic epidemiologic studies
  - Case control studies
  - Statistical testing
- Food testing

Epidemiologist use analytic epidemiology studies to search for determinants which can help explains the “why” and “how” component of a health event
Step 5: Solve point of contamination

Three types of data are used to link illnesses to contaminated foods and solve outbreaks:

**Epidemiologic data**
- Patterns in the time, place, or persons
- Food exposure information collected from patient interviews
- Cluster of unrelated sick people who ate at the same restaurant, shopped at the same grocery store, or attended the same event

**Traceback data**
- A traceback investigation starts with the sick people and works its way back through the chain of food production in search of a common point where contamination is likely to have occurred

**Food and environmental testing data**
- Collected by scientists looking for the pathogen that caused the illness in a food item collected from a sick patient’s home, a retail location, or in the food production environment
Step 6: Control an outbreak

Once a food source of an outbreak is implicated, control measures are taken

• If contaminated food stays on store shelves, in restaurant kitchens, or in home pantries, more people can get sick

• Control measures:
  • Cleaning and disinfecting food facilities
  • Temporary closing of restaurants of food processing plants
  • Recall of food items
  • Informing the public to avoid certain food products, or to throw out any suspect food products they may already have purchased
Step 7: Decide an outbreak is over

An outbreak is determined as concluded when the number of new illnesses drops back to expected levels

• Thanks to real time surveillance, we can monitor for on-going activity and respond accordingly

---

Number of People

Date of Illness Onset

Recall 1** Recall 2**
Bringing it all together

State Public Health Laboratories
Submit specimen and specimen data collected from sick patients, in real time

CDC
PulseNet National Database
Salmonella
Escherichia
Campylobacter
Listeria
Vibrio

Bioinformaticians
Query the database to link unrelated cases using WGS methods with the intent to identify emerging clusters of illness cases

Epidemiologist
- Collect, analyze, and interpret data from PulseNet
- Interview ill patients using hypothesis generating questionnaire
- Generate hypotheses to pinpoint source of contamination
- Test the hypothesis using analytic epidemiology studies
- Solve the point source contamination event
- Recommend control measures

Regulatory Agencies
- Help trace the identified common food source to the manufacturer and or distributor
- Take steps to identify lots of contamination foods to recall
- Can also go as far as to go to the alleged facilities to test machinery and equipment
- Make recommendations for food safety to the public

3/12/19
BIOL 7210 Computational Genomics
An example

<table>
<thead>
<tr>
<th>WGS_id</th>
<th>SourceState</th>
<th>PFGE-Xbal-pattern</th>
<th>PFGE-Blni-pattern</th>
<th>Outbreak</th>
<th>Age</th>
<th>SourceType</th>
<th>SourceSite</th>
<th>IsolatDate</th>
<th>ReceivedDat</th>
<th>UploadDate</th>
</tr>
</thead>
<tbody>
<tr>
<td>PNUSAE012639</td>
<td>PA</td>
<td>EXHX01.0047</td>
<td>EXHA26.0626</td>
<td>1804MLEXH-1</td>
<td>62</td>
<td>Human</td>
<td>Stool</td>
<td>2/8/18</td>
<td>2/26/18</td>
<td>3/12/18</td>
</tr>
<tr>
<td>PNUSAE012525</td>
<td>CA</td>
<td>EXHX01.0047</td>
<td>EXHA26.0626</td>
<td>1804MLEXH-1</td>
<td>71</td>
<td>Human</td>
<td>Stool</td>
<td>2/19/18</td>
<td>3/1/18</td>
<td>3/6/18</td>
</tr>
</tbody>
</table>
Sprinkle in some epidemiologic information

<table>
<thead>
<tr>
<th>WGS_id</th>
<th>Food consumed in the last 2 weeks</th>
</tr>
</thead>
<tbody>
<tr>
<td>PNUSAE012639</td>
<td>Apples, Salmon, Beans, Chicken, Bacon</td>
</tr>
<tr>
<td>PNUSAE012525</td>
<td>Turkey, Carrots, Tuna, Cereal, Eggs</td>
</tr>
<tr>
<td>PNUSAE013041</td>
<td>Chicken, Cheese, Tomato, Soup, Frozen food</td>
</tr>
<tr>
<td>PNUSAE013098</td>
<td>Pizza, Apples, Canteloupe, Frozen food, Venison</td>
</tr>
<tr>
<td>PNUSAE013124</td>
<td>Frozen food, Sausage, Bread, Chicken, Eggs</td>
</tr>
<tr>
<td>WGS_id</td>
<td>Food consumed in the last 2 weeks</td>
</tr>
<tr>
<td>--------------</td>
<td>--------------------------------------------</td>
</tr>
</tbody>
</table>
| PNUSAE012639 | Apples  
  Salmon  
  Beans  
  Chicken  
  Bacon                                        |
| PNUSAE012525 | Turkey  
  Carrots  
  Tuna  
  Cereal  
  Eggs                                          |
| PNUSAE013041 | Chicken  
  Cheese  
  Tomato  
  Soup  
  Frozen food                                   |
| PNUSAE013098 | Pizza  
  Apples  
  Canteloupe  
  Frozen food  
  Venison                                       |
| PNUSAE013124 | Frozen food  
  Sausage  
  Bread  
  Chicken  
  Eggs                                          |

<table>
<thead>
<tr>
<th>WGS_id</th>
<th>Food consumed in the last 2 weeks</th>
</tr>
</thead>
</table>
| PNUSAE012639 | Apples  
  Salmon  
  Beans  
  Chicken  
  Bacon                                        |
| PNUSAE012525 | Turkey  
  Carrots  
  Tuna  
  Cereal  
  Eggs                                          |
| PNUSAE013041 | Chicken  
  Cheese  
  Tomato  
  Soup  
  Frozen food                                   |
| PNUSAE013098 | Pizza  
  Apples  
  Canteloupe  
  Frozen food  
  Venison                                       |
| PNUSAE013124 | Frozen food  
  Sausage  
  Bread  
  Chicken  
  Eggs                                          |

<table>
<thead>
<tr>
<th>WGS_id</th>
<th>Food consumed in the last 2 weeks</th>
</tr>
</thead>
</table>
| PNUSAE012639 | Apples  
  Salmon  
  Beans  
  Chicken  
  Bacon                                        |
| PNUSAE012525 | Turkey  
  Carrots  
  Tuna  
  Cereal  
  Eggs                                          |
| PNUSAE013041 | Chicken  
  Cheese  
  Tomato  
  Soup  
  Frozen food                                   |
| PNUSAE013098 | Pizza  
  Apples  
  Canteloupe  
  Frozen food  
  Venison                                       |
| PNUSAE013124 | Frozen food  
  Sausage  
  Bread  
  Chicken  
  Eggs                                          |
Enter genome scale resolution
Remove the noise

### Table

<table>
<thead>
<tr>
<th>WGS_id</th>
<th>SourceState</th>
<th>PFGE-Xbal-pattern</th>
<th>PFGE-BlnI-pattern</th>
<th>Outbreak</th>
<th>Age</th>
<th>SourceType</th>
<th>SourceSite</th>
<th>IsolatDate</th>
<th>ReceivedDate</th>
<th>UploadDate</th>
</tr>
</thead>
<tbody>
<tr>
<td>PNUSA012639</td>
<td>PA</td>
<td>EXHX01.0047</td>
<td>EXHA26.0626</td>
<td>1804MLEXH-1</td>
<td>62</td>
<td>Human</td>
<td>Stool</td>
<td>2/8/18</td>
<td>2/26/18</td>
<td>3/12/18</td>
</tr>
<tr>
<td>PNUSA012525</td>
<td>CA</td>
<td>EXHX01.0047</td>
<td>EXHA26.0626</td>
<td>1804MLEXH-1</td>
<td>71</td>
<td>Human</td>
<td>Stool</td>
<td>2/19/18</td>
<td>3/1/18</td>
<td>3/6/18</td>
</tr>
</tbody>
</table>

### Food consumed in the last 2 weeks

- Apples
- Salmon
- Beans
- Chicken
- Bacon
- Turkey
- Carrots
- Tuna
- Cereal
- Eggs
- Chicken
- Cheese
- Tomato
- Soup
- Frozen food
- Pizza
- Apples
- Cantaloupe
- Frozen food
- Venison
- Frozen food
- Sausage
- Bread
- Chicken
- Eggs

3/12/19
BIOL 7210 Computational Genomics
40
Supporting genome-based results

<table>
<thead>
<tr>
<th></th>
<th>PNUSAE012525</th>
<th>PNUSAE012639</th>
<th>PNUSAE013041</th>
<th>PNUSAE013098</th>
<th>PNUSAE013124</th>
</tr>
</thead>
<tbody>
<tr>
<td>PNUSAE012525</td>
<td>0</td>
<td>15</td>
<td>17</td>
<td>17</td>
<td>21</td>
</tr>
<tr>
<td>PNUSAE012639</td>
<td>15</td>
<td>0</td>
<td>12</td>
<td>12</td>
<td>13</td>
</tr>
<tr>
<td>PNUSAE013041</td>
<td>17</td>
<td>12</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>PNUSAE013098</td>
<td>17</td>
<td>12</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>PNUSAE013124</td>
<td>21</td>
<td>13</td>
<td>1</td>
<td>1</td>
<td>0</td>
</tr>
</tbody>
</table>

**Virulence determinants**

- stx1a, stx2d, eae var. 2, ehxA var. 11
- stx1a, stx2d, eae var. 2, ehxA var. 11
- stx1a, stx2d, eae var. 2, ehxA var. 11
- stx1a, stx2g, eae var. 4, ehxA var. 3
- stx1c, stx2f, eae var. 4, ehxA var. 5
Summary: epidemiological context

Epidemiologists utilize patient information and genomic analysis data to conduct investigations into foodborne illness outbreaks to identify sources of contamination.

Bioinformatics scientists provide sharpened tools so that epidemiologist can conduct their investigations more effectively.

The ultimate mission of the CDC is to protect the public against expensive and dangerous health threats.
References
