Decision Tree Analysis to Predict Food Sources of Shiga Toxin-Producing Escherichia coli (STEC) Outbreaks using Demographic and Outbreak Characteristics, United States, 1998-2014

Alice White, MS

Colorado Integrated Food Safety Center of Excellence

CSTE Annual Conference, Boise ID, June 2017
Integrated Food Safety Centers of Excellence (CoE)
CoE Mission

Build capacity in local and state health departments by developing and providing online and in-person resources, training, and assistance for foodborne illness surveillance and investigations.
Unsolved Outbreaks

• Outbreak investigations identify food safety interventions
  - Pathogen, food, contributing factors

• Identifying food vehicle is critical
  - 42% of outbreaks do not confirm food vehicle
### Demographics – STEC O157 Outbreak

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Age, median (range)</strong></td>
<td>24 (4-66)</td>
</tr>
<tr>
<td><strong>Gender, n (%)</strong></td>
<td></td>
</tr>
<tr>
<td>Female</td>
<td>21 (64%)</td>
</tr>
<tr>
<td>Male</td>
<td>12 (36%)</td>
</tr>
</tbody>
</table>
Project Purpose

• Use existing outbreak data to develop methods and tools to make hypotheses about food vehicles

• Used STEC outbreaks to describe differences in demographic and outbreak characteristics and predict food vehicles using differential characteristics
Methods

• Data source
  – Electronic Foodborne Outbreak Reporting System (eFORS), 1998-2008
  – National Outbreak Reporting System (NORS), 2009-2014

• Food source categories
  – Interagency Food Safety Analytics Collaboration (IFSAC) Food Categorization Scheme

• Predictors
  – Percent female, percent hospitalized, percent age categories (< 5 years, 5-19, 20-49, >50), multistate exposure, setting (private or non-private), season, duration
Results

STEC Outbreaks, 1998-2014 (n=470)

Undetermined food (n=153)
Complex foods (n=80)
Rare foods (n=31)
Non-O157:H7 (n=15)

119 BEEF

25 DAIRY

39 LEAFY GREENS
Univariate Results

<table>
<thead>
<tr>
<th>Predictor</th>
<th>Beef</th>
<th>Dairy</th>
<th>Leafy</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td>**Gender *</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>% Female</td>
<td>50 (34-70)</td>
<td>50 (33-67)</td>
<td>64 (56-74)</td>
<td>0.009</td>
</tr>
<tr>
<td>**Age (years) *</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>% &lt;5</td>
<td>0 (0-13)</td>
<td>22 (0-33)</td>
<td>0 (0-4)</td>
<td>0.014</td>
</tr>
<tr>
<td>% 5-19</td>
<td>35 (12-61)</td>
<td>50 (50-67)</td>
<td>20 (15-33)</td>
<td>0.017</td>
</tr>
<tr>
<td>% 20-49</td>
<td>21 (0-33)</td>
<td>17 (0-33)</td>
<td>44 (30-54)</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>% &gt;50</td>
<td>14 (0-33)</td>
<td>0 (0-0)</td>
<td>19 (9-40)</td>
<td>0.002</td>
</tr>
<tr>
<td>**Cases *</td>
<td>9 (3-17)</td>
<td>5 (3-9)</td>
<td>18 (10-33)</td>
<td>&lt;0.001</td>
</tr>
</tbody>
</table>

* Data presented as median (IQR), p value from Kruskal-Wallis
## Univariate Results

<table>
<thead>
<tr>
<th>Predictor</th>
<th>Beef</th>
<th>Dairy</th>
<th>Leafy</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Setting†</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Private Home</td>
<td>45 (52%)</td>
<td>14 (70%)</td>
<td>5 (13%)</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Non-private</td>
<td>33 (38%)</td>
<td>4 (20%)</td>
<td>32 (82%)</td>
<td></td>
</tr>
<tr>
<td><strong>Multistate Exposure†</strong></td>
<td>31 (36%)</td>
<td>2 (10%)</td>
<td>22 (56%)</td>
<td>0.002</td>
</tr>
<tr>
<td><strong>Season‡</strong></td>
<td></td>
<td></td>
<td></td>
<td>0.108</td>
</tr>
<tr>
<td>Fall</td>
<td>17 (20%)</td>
<td>7 (35%)</td>
<td>17 (44%)</td>
<td></td>
</tr>
<tr>
<td>Spring</td>
<td>28 (33%)</td>
<td>6 (30%)</td>
<td>8 (21%)</td>
<td></td>
</tr>
<tr>
<td>Summer</td>
<td>34 (40%)</td>
<td>6 (30%)</td>
<td>9 (23%)</td>
<td></td>
</tr>
<tr>
<td>Winter</td>
<td>7 (8%)</td>
<td>1 (5%)</td>
<td>5 (13%)</td>
<td></td>
</tr>
</tbody>
</table>

† Presented as proportion (%), p value from Chi-square
‡ Presented as proportion (%), Fisher’s Exact
Food Source Prediction of Shiga Toxin–Producing *Escherichia coli* Outbreaks Using Demographic and Outbreak Characteristic, United States, 1998–2014

Alice White¹, Alicia Cronquist², Edward J. Bedrick³, and Elaine Scallan¹
STEC O157:H7 outbreak

Cases 20-49 years < 33%?

- yes
  - Cases under 5 years < 10%?
    - yes
      - Is the season fall?
        - yes
          - Female cases < 63%?
            - yes
              - BEEF
            - no
              - LEAFY GREENS
        - no
          - BEEF
    - no
      - Cases under 5 years < 68%?
        - yes
          - BEEF
        - no
          - DAIRY

- no
  - Cases under 5 years < 17%?
    - yes
      - BEEF
    - no
      - LEAFY GREENS
STEC O157:H7 outbreak

Cases 20-49 years < 33%?

yes

Cases under 5 years < 10%?

yes

Is the season fall?

yes

Female cases < 63%?

yes

BEEF

no

no

BEEF

no

no

Cases under 5 years < 17%?

yes

Cases under 5 years < 68%?

yes

Cases under 5 years < 39%?

no

no

BEEF

DAIRY

DAIRY

LEAFY GREENS
## Decision Tree Results

<table>
<thead>
<tr>
<th>PREDICTED OUTCOME</th>
<th>OBSERVED OUTCOME</th>
<th>Beef</th>
<th>Dairy</th>
<th>Leafy</th>
<th>Overall</th>
<th>Precision (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Beef</td>
<td>Beef</td>
<td>109</td>
<td>13</td>
<td>19</td>
<td>141</td>
<td>77</td>
</tr>
<tr>
<td></td>
<td>Dairy</td>
<td>4</td>
<td>10</td>
<td>1</td>
<td>15</td>
<td>67</td>
</tr>
<tr>
<td></td>
<td>Leafy</td>
<td>6</td>
<td>2</td>
<td>19</td>
<td>27</td>
<td>70</td>
</tr>
<tr>
<td></td>
<td>Overall</td>
<td>119</td>
<td>25</td>
<td>39</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sensitivity (%)</td>
<td>Beef</td>
<td>92</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Dairy</td>
<td>40</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Leafy</td>
<td>49</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Specificity (%)</td>
<td>Beef</td>
<td>50</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Dairy</td>
<td>97</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Leafy</td>
<td>94</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Discussion

• Pathogen & serotype are the best predictors of a food vehicle
• Demographic data (age & gender) and season can provide some discrimination between vehicles
Discussion

• Limitations
  – Known sources of STEC illness

• Future directions
  – Prospective validation
  – Develop tools
  – Other major pathogens
Acknowledgments

• Co-authors
  – Kayla Bell, Colorado School of Public Health
  – Elaine Scallan, PhD, Colorado School of Public Health

• NORS team at CDC
www.COFoodSafety.org
alice.white@ucdenver.edu
(303) 724-9319