Distribution Agreement

In presenting this thesis or dissertation as a partial fulfillment of the requirements for an advanced degree from Emory University, I hereby grant to Emory University and its agents the non-exclusive license to archive, make accessible, and display my thesis or dissertation in whole or in part in all forms of media, now or hereafter known, including display on the world wide web. I understand that I may select some access restrictions as part of the online submission of this thesis or dissertation. I retain all ownership rights to the copyright of the thesis or dissertation. I also retain the right to use in future works (such as articles or books) all or part of this thesis or dissertation.

Signature:

____ _______________

Paula D. Strassle Date

Age and Gender Distributions of Persons in Foodborne Disease Outbreaks and Associations with Food Commodities

By

Paula D. Strassle

MSPH

Epidemiology

Dr. L. Hannah Gould

Committee Chair

Dr. John McGowan

Committee Chair

Age and Gender Distributions of Persons in Foodborne Disease Outbreaks and Associations with Food Commodities

By

Paula D. Strassle

Bachelor of Science

University of Maryland, Baltimore County

2010

Thesis Committee Chairs: Dr. L. Hannah Gould, PhD and Dr. John McGowan, MD

An abstract of

A thesis submitted to the Faculty of the
Rollins School of Public Health of Emory University

in partial fulfillment of the requirements for the degree of
Master of Science in Public Health
in Epidemiology

2014

Abstract

Age and Gender Distributions of Persons in Foodborne Disease Outbreaks and Associations to Contaminated Food Source

By Paula D. Strassle

Background. Foodborne disease accounts for 9.4 million illnesses caused by known pathogens in the United States every year. Identifying outbreak sources are a critical part of containment and removal. There is the potential to use demographic information to assist in the prediction and identification of contamination sources in foodborne disease outbreaks. This analysis assesses the association of gender and age distributions of persons in foodborne disease outbreaks with specific commodity sources. *Methods.* Outbreak data from January 1, 1998 to December 31, 2011 from the CDC Foodborne disease outbreak Surveillance System (FDOSS). Crude and adjusted odds-ratios were calculated for the seventeen defined commodities and a classification tree based-model was constructed to assess the predictability of demographic factors. *Results.* Age and gender were associated with multiple commodities after adjustment; men were more likely to be involved in outbreaks caused by mollusks (aOR=2.41, 95%CI 1.92, 3.03), dairy (aOR=1.60, 95%CI 1.31, 1.95), beef (aOR=1.43, 95%CI 1.24, 1.64), game (aOR=7.66, 95%CI 3.68, 15.94), and pork (aOR=1.46, 95%CI 1.23, 1.72). Women were more likely to be involved in outbreaks caused by fruit-nuts (aOR=2.61, 95%CI 2.13, 3.20), leafy vegetables (aOR=1.85, 95%CI 1.57, 2.18), sprouts (aOR=3.35, 95%CI 1.97, 5.70), and vine-stalk vegetables (aOR=1.53, 95%CI 1.14, 2.05). Higher proportions of children (under 4 and 5-19), were involved in outbreaks caused by dairy, aOR=4.10 (95%CI 3.37, 5.00) and 3.47 (95%CI 2.85, 4.23), respectively, beef, aOR=1.50 (95%CI 1.27, 1.78) and 1.51 (95%CI 1.31, 1.73), pork, aOR=1.46 (95%CI 1.20, 1.78), and 1.45 (95%CI 1.23, 1.71) and poultry, aOR=1.30 (95%CI 1.13, 1.50) and 1.24 (95%CI 1.11, 1.40). Older adults were more commonly involved in outbreaks caused by fish (aOR=1.54, 95%CI 1.29, 1.83), mollusks (aOR=1.62, 95%CI 1.23, 2.15), fruit-nuts (aOR=2.78, 95%CI 1.24, 6.23), root vegetables (aOR=2.18, 95%CI 1.45, 3.28), and vine-stalk vegetables (aOR=1.68, 95%CI 1.20, 2.36). The tree based model had a kappa agreement of 0.17. *Discussion.* This analysis provides preliminary support that age and gender groups are more likely to be associated with certain foods causing outbreaks. Further analysis is needed in to better understand these relationships and the utility of demographic profiles to help in predicting food sources in outbreaks.

Age and Gender Distributions of Persons in Foodborne Disease Outbreaks and Associations to Contaminated Food Source

By

Paula D. Strassle

Bachelor of Science

University of Maryland, Baltimore County

2010

Thesis Committee Chairs: Dr. L. Hannah Gould, PhD and Dr. John McGowan, MD

A thesis submitted to the Faculty of the

Rollins School of Public Health of Emory University

in partial fulfillment of the requirements for the degree of
Master of Science in Public Health
in Epidemiology

2014

Table of Contents

Introduction

It is estimated that foodborne disease accounts for 9.4 million illnesses caused by known pathogens in the United States every year (1). Although most foodborne illnesses are sporadic, outbreak-associated cases are an important source of information about the foods and pathogens causing illness (2). Identifying the source of an outbreak is a critical part of containment and removal of the contaminated food source. Typically, outbreak investigations initially consist of identifying cases and interviewing them to determine potential sources of contamination or transmission. These hypothesis-generating interviews contain extensive lists of food items and attempt to find commonalities in consumption patterns or locations between cases (5, 6). It is sometimes necessary to conduct multiple interviews with each case before potential sources can be identified. Moreover, as new food vehicles are more frequently identified, for example the recent listeriosis outbreak caused by cantaloupes, targeting potential food sources based on pathogen may become less effective in outbreak investigations (8).

There is the potential to utilize demographic information from outbreak cases to assist in the identification of contaminated food sources. Several factors can affect the quantity and variety of food an individual or family consumes, including income, gender, age, and health status (9-17). Gender, age, and state of residence based differences in food consumption have been observed using national survey data within the United States (9-17). However, there are no studies examining demographics of persons involved in foodborne disease outbreaks and whether there are differences by implicated food.

To address this need for further information, this thesis will analyze the demographic distributions of foodborne disease outbreak cases, specifically age and gender, to determine if observed food consumption patterns exist. It is hypothesized that gender and age will be associated with specific food commodities; specifically that men will be more likely to be involved in outbreaks caused by meat, women will be more likely to be involved in outbreaks caused by produce, and children will be more likely to be involved in outbreaks caused by dairy and fruit. Moreover, demographic data will be used to create a classification tree based model to determine the predictability of demographic distributions on specific contaminated commodities.

With outbreaks occurring in novel sources, targeted questionnaires based on pathogens are becoming less applicable and open-ended interviews are time consuming. There is a clear gap of knowledge concerning demographic trends in foodborne disease, and research in this area could provide novel and useful ways to investigate foodborne disease outbreaks in the United States.

Literature Review

Foodborne Illness Surveillance

It is estimated that foodborne disease accounts for 9.4 million illnesses caused by known pathogens in the United States every year (1). Although most foodborne illnesses are sporadic, outbreak-associated illnesses are an important source of information about the foods and pathogens causing illness (2). In 2009-2010 there were a reported 1,527 foodborne disease outbreaks in the United States, accounting for over 29,000 illnesses, 1,184 hospitalizations, and 23 deaths (3). Of these 1,527 outbreaks, 299 (20%) were attributed to a single food commodity (3). Identifying the source of an outbreak is a critical part of containment and removal of the contaminated food source.

The Centers for Disease Control and Prevention (CDC) monitors and evaluates foodborne disease outbreaks via the Foodborne Disease Outbreak Surveillance System (FDOSS). FDOSS is a passive national surveillance system with state, local, territorial, and national public health agencies reporting foodborne disease outbreaks caused by bacterial, viral, parasitic, and chemical agent since 1973 (2,3). While only a minority of illnesses, hospitalizations and deaths from foodborne illness are reported to health departments, and therefore FDOSS, this system provides valuable information on the pathogenic agents that cause foodborne diseases, the implicated foods in outbreaks, and the settings in which contamination and transmission occur (2, 3). To better assist the reporting of outbreaks, regulation of food safety, and increase understanding of implicated food sources, a scheme to categorize foods reported in outbreaks was created in 2009 (4).

Foodborne disease outbreak Investigations

Typically, outbreak investigations consist of identifying cases and interviewing them to determine potential sources of contamination or transmission. These hypothesis-generating interviews review extensive lists of food items and attempt to find commonalities in consumption patterns or locations between cases (5, 6). Currently, hypothesis-generating interviews are one of the primary methods used in outbreak investigations, even though this technique possesses several limitations and is time-consuming (5). First, these interviews are highly dependent on the memories of patients; this is further expounded by the typical lag of weeks between initial exposure and the start of the outbreak investigation. Additionally, when the contaminated food is an ingredient in a larger dish (e.g., eggs, spices, or herbs), cases may not even know that they consumed it (5). It is sometimes necessary to conduct multiple interviews with each case before potential sources can be identified.

In 2009, a standardized targeted questionnaire- focusing on foods associated with previous outbreaks of *E. coli* proved ineffective at hypothesis generation in a multi-state outbreak of *Escherichia coli* O157:H7 infections (7). After completing the first set of interviews, a second round of open-ended interviews with patients was performed. Based on the secondary interviews, a second questionnaire was created for a matched case-control study. Through this analysis, prepackaged cookie dough was determined to be the source of the outbreak. Prepackaged cookie dough had previously never been implicated in a foodborne disease outbreak and therefore was not included in the first questionnaire (7). Similarly, in a 2011 nation-wide listeriosis outbreak, cantaloupe was determined to be the vehicle of transmission, even though produce is rarely contaminated by this pathogen (8). As these novel outbreak vehicles become more common, targeting potential food sources based on pathogen may become less effective in outbreak investigations.

<u>Food Consumption Patterns</u>

Several factors can affect the quantity and variety of food an individual or family consumes, including income, gender, age, and health status (9-17). Most of the research assessing food consumption patterns has occurred outside of the United States, and is therefore not directly applicable (9, 10, 12). However, gender and age based differences have been recently observed in food consumption using national survey data (9-17).

The Foodborne Diseases Active Surveillance Network (FoodNet) is a collaborative program with the CDC, United States Department of Agriculture, and Food and Drug Administration (FDA). FoodNet conducts periodic surveys of the general population of 10 states, via random digit dialing, to create a disproportionate stratified sample representative of the US population. Questions on the survey include questions about diarrheal illness, demographics, and potential exposures (9, 11). Potential exposure to 'risky foods' includes the consumption of pink or undercooked chicken, turkey, hamburgers or ground pork, raw fish or shellfish, unpasteurized milk, runny eggs, alfalfa sprouts, and unpasteurized apple juice (9,11).

Men have been found more likely to consume 'high risk' foods, including unpasteurized milk, raw shellfish, runny eggs, and pink hamburgers when compared with women (9-12). In an analysis of the 1996-1997 survey data (which included residents of California, Connecticut, Georgia, Minnesota, and Oregon) age, gender, and state of residence were found to impact the preferences high-risk foods (9). In an analysis of the date collected during 1998-1999, which was conducted in 7 states (same as above, plus Maryland and New York) found similar results (10). Men were statistically more likely to report consuming at least one risky food when compared with women, regardless of age (9, 10).

Shiferaw *et al.* (2012) also found similar results to the prior studies when analyzing the 2006-2007 FoodNet population survey data but looked at both commodity categories and specific foods (11). In this analysis, women reported eating more fruits- including apples, strawberries, raspberries, blueberries, blackberries, and cantaloupe (11). There were also significant differences in the proportion of men and women eating specific vegetables, with women consuming more celery, carrots, cucumbers, zucchini (or other 'soft squash'), avocadoes, and tomatoes. Men reported consuming more corn and Brussels sprouts (11). Additionally, men consumed more steak, game birds, ham, oysters, shrimp, and nuts (11). In regards to high-risk foods, men had a statistically significant higher rate of consuming undercooked hamburger, raw oysters, and runny eggs; women were more likely to eat alfalfa sprouts (11). In a study on meat consumption trends in the United States, men consumed more meat (of every variety) than women (12).

In a study of school children and adolescents, gender also was found to play a role in the reported consumption and preference for certain foods (14). In a 5-point rating scale survey, girls had a higher preference for fruits and vegetables. Boys reported having a higher preference for fish, beef, and pork (14). Moreover, these preferences were associated with school level, indicating that these differences in preference are also impacted by age.

Additionally age has also been associated with differences in food consumption patterns. The proportion of young adults that reported consuming raw shellfish, runny eggs and alfalfa sprouts is higher than older adults (10). Additionally, older adults are likely to eat risky foods then younger age groups (11). Meat consumption, with the exception of fish, occurred highest in adults; peak fish consumption occurred in the older adult age group. Children aged 2-11 were least likely to consume meat when compared with persons in the other age groups (13).

The National Health and Nutrition Examination Survey (NHANES), a continuous nationally representative cross-sectional survey, was analyzed to assess fruit and vegetable consumption among adolescents aged 12-18, and adults (15). While this study also focused on health and meeting daily recommendations for fruit and vegetable intake, the study concluded that adolescents were less likely to report consuming raw fruits when compared with adults (15). Dark green vegetables were consumed by adult women more often than adolescents; however, adolescents were more likely to report consuming tomatoes. In general, adult women were more likely to eat raw fruits and vegetables than adult men and adolescents of either gender (15).

There is the potential to use consumption patterns to better predict and assist in the identification of contamination sources in foodborne disease outbreaks. In a 2004 *Salmonella enterica* (serotype Enteritidis) outbreak in Oregon, hypothesis-generating interviews proved unsuccessful in identifying a common exposure among cases (18). However, in 2002 FoodNet performed a population survey, which included residents of Oregon (19). Using a binomial distribution, the background rates of food consumption of residents of Oregon was compared with consumption estimates for cases in the outbreak (18). While the population survey estimated that only 9% of Oregon residents consume raw almonds any given week, all 5 of the cases in the outbreak reported eating almonds within 5 days of becoming sick; this information was used to ultimately help identify almonds as the source of the outbreak (18). Similarly, in 2008, the source of an outbreak of *Salmonella enterica* serotype Agona infections was determined by comparing the percentage of cases (84%) that reported eating puffed rice cereal with the total ready-to-eat cereal market share in the US (0.063%) (20).

Food Consumption Patterns in Outbreaks

Anecdotally, evidence of age and gender patterns has been observed in foodborne disease outbreaks. For example, in the raw cookie dough outbreak discussed above, 66% of cases were under the age of 19 and 71% were female. Overall, women 19 years old or younger made up 55% of the total cases (7). Out of the 147 cases in the listeriosis cantaloupe outbreak, 85 were women (58%) and 127 were 60 years or older (86%) (8). In the two recently reported milk outbreaks, over 30% of cases in each were in children (21, 22). Using the FDOSS reporting system and the food commodity breakdown instituted by the CDC, there is the potential to analyze demographic trends in foodborne disease outbreaks as it relates to pathogen and food vehicles.

There is a clear gap in knowledge concerning the foods associated with persons who have illness by demographics. Research in this area could provide novel and useful ways to investigate foodborne disease outbreaks in the United States. Using data from foodborne disease outbreaks, the goal of this thesis is to determine if gender and age are associated, and potentially predictive, of contaminated food sources in foodborne disease outbreaks.

References

1. Scallan E, Hoekstra RM, Angulo FJ, Tauxe RV, Widdowson MA, Roy SL, Jones JL, Griffin PM. Foodborne illnesses acquired in the United States- major pathogens. *Emerg Infect Dis*. 2011; 17(1): 7-15.

2. Gould LH, Walsh KA, Vieira AR, Herman K, Williams IT, Hall AJ and Cole D. Surveillance for foodborne disease outbreaks—United States, 1998-2008. *MMWR Surveill Summ*. 2013; 62(2):1-34.

3. Centers for Disease Control and Prevention (CDC). Surveillance for foodborne disease outbreaks—United States, 2009-2010. *MMWR Morb Mortal Wkly Rep*. 2013; 62(3):41-47.

4. Painter JA, Ayers T, Woodruff R, Blanton E, Perez N, Hoekstra RM, Griffin PM and Braden C. Recipes for foodborne outbreaks: A scheme for categorizing and grouping implicated foods. *Foodborne Pathog Dis*. 2009; 6(10): 1259-1264.

5. Centers for Disease Control and Prevention. Generating hypotheses about likely sources. http://www.cdc.gov/foodsafety/outbreaks/investigating-outbreaks/investigations/sources.html. Updated November 12, 2013. Accessed December 2, 2013.

6. Pires SM, Evers EG, van Pelt W, Ayers T, Scallan E, Angulo FJ, Havelaar, Hald T, Med-Vet-Net Workpackage 28 Working Group. Attributing the human disease burden of foodborne infections to specific sources. *Foodborne Pathog Dis*.

7. Neil KP, Biggerstaff G, MacDonald JK, Trees E, Medus C, Musser KA, Stroika SG, Zink D and Sotir MJ. A novel vehicle for transmission of *Escherichia coli* O157:H7 to humans: Multistate outbreak of *E. coli* O157:H7 infections associated with consumption of ready-to-break commercial prepackaged cookie dough-United States, 2009. *Clin Infect Dis*. 2012; 54(4): 511-518.

8. McCollum JT, Cronquist AB, Silk BJ, et al. Multistate outbreak of Listeriosis associated with cantaloupe. *N Engl J Med*. 2013; 369(10): 944-953.

9. Shiferaw B, Yang S, Cieslak P, Vugia D, Marcus R, Koehler J, Deneen V, Angulo F, the FoodNet Working Group. Prevalence of high-risk food consumption and food-handling practices among adults: A multi-states survey, 1996 to 1997. *J Food Protection*. 2000; 63(11): 1538-1543.

10. Samuel MC, Vugia DJ, Koehler KM, Marcus R, Deneen V, Damaske B, Shiferaw B, Hadler J, Henao OL, Angulo FJ. Consumption of risky foods among adults at high risk for severe foodborne diseases: Room for improved targeted prevention messages. *J Food Safety*. 2007; 27: 219-232.

11. Shiferaw B, Verrill L, Booth H, Zansky SM, Norton DM, Crim S, Henao OL. Sex-based differences in food consumption: Foodborne diseases active surveillance network (FoodNet) population survey, 2006-2007. *Clin Infect Dis*. 2012; 54 Suppl 5: S453-S457.

12. Daniel CR, Cross AJ, Koebnick C, Sinha R. Trends in meat consumption in the USA. *Public Health Nutr*. 2011; 14(4): 575-583.

13. Wardle J, Haase AM, Steptoe A, Nillapun M, Jonwutiwes K, Bellisle F. Gender differences in food choice: the contribution of health beliefs and dieting. *Ann Behav Med*. 2004; 27(2): 107-116.

14. Caine-Bish NL, Scheule B. Gender differences in food preferences of school-aged children and adolescents. *J Sch Health*. 2009; 79(11): 532-540.

15. Kimmons J, Gillespie C, Seymour J, Serdula M, Blanck HM. Fruit and vegetable intake among adolescents and adults in the United States: Percentage meeting individualized recommendations. *Medscape J Med*. 2009; 11(1): 26.

16. Emanuel AS, McCully SN, Gallagher KM, Updegraff JA. Theory of planned behavior explains gender difference in fruit and vegetable consumption. *Appetite*. 2012; 59 693-697.

17. Dehghan M, Akhtar-Danesh N, Merchant AT. Factors associated with fruit and vegetable consumption among adults. *J Hum Nutr Diet*. 2011; 24(2): 128-134.

18. Centers for Disease Control and Prevention (CDC). Outbreak of Salmonella serotype Enteritidis infections associated with raw almonds—United States and Canada, 2003-2004. *MMWR Morb Mortal Wkly Rep*. 2004; 53(22): 484-487.

19. Centers for Disease Control and Prevention. Foodborne Diseases Active Surveillance Network (FoodNet): Population survey atlas of exposures, 2002. Atlanta, Georgia: US Department of Health and Human Resources, CDC, 2004.

20. Russo ET, Biggerstaff G, Hoekstra RM, Meyer S, Patel N, Miller B, Quick R. A recurrent, multistate outbreak of *Salmonella* serotype Agona infections associated with dry, unsweetened cereal consumption, United States, 2008. *J Food Protect*. 2013; 76(2): 227-230.

21. Longenberger AH, Palumbo AJ, Chu AK, Moll ME, Weltman A, Ostroff SM. *Campylobacter jejuni* infections associated with unpasteurized milk- multiple states, 2012. *Clin Infect Dis*. 2013; 57(2): 263-266.

22. Centers for Disease Control and Prevention (CDC). Recurrent outbreak of *Campylobacter jejuni* infections associated with raw milk dairy- Pennsylvania, April-May 2013. *MMWR Morb Mortal Wkly Rep*. 2013; 62(34):702.

Age and Gender Distributions of Persons in Foodborne Disease Outbreaks and Associations with Food Commodities

P. Strassle[1], W. Gu[2] and L. H. Gould[2]

[1]Epidemiology Department, Rollins School of Public Health, Emory University

[2]Division of Foodborne, Waterborne, and Environmental Diseases, Centers for Disease Control and Prevention

Abstract

Background. It is estimated that foodborne disease accounts for 9.4 million illnesses caused by known pathogens in the United States every year; identifying the source of an outbreak is a critical part of containment and removal of the contaminated food source. There is the potential to use demographic information to assist in the prediction and identification of contamination sources in foodborne disease outbreaks. This analysis assesses the association of gender and age distributions of persons in foodborne disease outbreaks with specific commodity sources. *Methods*. Outbreak data from January 1, 1998 to December 31, 2011 from the CDC Foodborne disease outbreak Surveillance System (FDOSS). Crude and adjusted odds-ratios were calculated for the seventeen defined commodities and a classification tree based-model was constructed to assess the predictability of demographic factors. *Results*. Age and gender were associated with multiple commodities after adjustment; men were more likely to be involved in outbreaks caused by mollusks (aOR=2.41, 95% CI 1.92, 3.03), dairy (aOR=1.60, 95% CI 1.31, 1.95), beef (aOR=1.43, 95% CI 1.24, 1.64), game (aOR=7.66, 95% CI 3.68, 15.94), and pork (aOR=1.46, 95% CI 1.23, 1.72). Women were more likely to be involved in outbreaks caused by fruit-nuts (aOR=2.61, 95% CI 2.13, 3.20), leafy vegetables (aOR=1.85, 95% CI 1.57, 2.18), sprouts (aOR=3.35, 95% CI 1.97, 5.70), and vine-stalk vegetables (aOR=1.53, 95% CI 1.14, 2.05). Higher proportions of children (under 4 and 5-

19), were involved in outbreaks caused by dairy, aOR=4.10 (95% CI 3.37, 5.00) and 3.47 (95% CI 2.85, 4.23), respectively, beef, aOR=1.50 (95% CI 1.27, 1.78) and 1.51 (95% CI 1.31, 1.73), pork, aOR=1.46 (95% CI 1.20, 1.78), and 1.45 (95% CI 1.23, 1.71) and poultry, aOR=1.30 (95% CI 1.13, 1.50) and 1.24 (95% CI 1.11, 1.40). Older adults were more commonly involved in outbreaks caused by fish (aOR=1.54, 95% CI 1.29, 1.83), mollusks (aOR=1.62, 95% CI 1.23, 2.15), fruit-nuts (aOR=2.78, 95% CI 1.24, 6.23), root vegetables (aOR=2.18, 95% CI 1.45, 3.28), and vine-stalk vegetables (aOR=1.68, 95% CI 1.20, 2.36). The tree based model had a kappa agreement of 0.17. *Discussion*. This analysis provides preliminary support that age and gender groups are more likely to be associated with certain foods causing outbreaks. Further analysis is needed in to better understand these relationships and the utility of demographic profiles to help in predicting food sources in outbreaks.

Background

It is estimated that foodborne disease accounts for 9.4 million illnesses caused by known pathogens in the United States every year (1). Although most foodborne illnesses are sporadic, outbreak-associated illnesses are an important source of information about the foods and pathogens causing illness (2). In 2009-2010 there were a reported 1,527 foodborne disease outbreaks in the United States, accounting for over 29,000 illnesses, 1,184 hospitalizations, and 23 deaths (3). Of these 1,527 outbreaks, 299 (20%) were attributed to a single food commodity (3). Identifying the source of an outbreak is a critical part of containment and removal of the contaminated food source.

Typically, outbreak investigations initially consist of identifying cases and interviewing them to determine potential sources of contamination or transmission. These hypothesis-generating interviews review extensive lists of food items and attempt to find commonalities in consumption patterns or locations between cases (4, 5).

Currently, hypothesis-generating interviews are one of the primary methods used in outbreak investigations, even though this technique possesses several limitations and is time-consuming (4). First, these interviews are highly dependent on the memories of patients; this is further expounded by the typical lag of weeks between initial exposure and the start of the outbreak investigation. Additionally, when the contaminated food is an ingredient in a larger dish (e.g. eggs, spices, or herbs), cases may not even know that they consumed it (4). It is sometimes necessary to conduct multiple interviews with each case before potential sources can be identified.

In 2009, a standardized targeted questionnaire- focusing on foods associated with previous outbreaks of *E. coli* proved ineffective at hypothesis generation in a multi-state outbreak of *Escherichia coli* O157:H7 infections (6). After completing the first set of interviews, a second round of open-ended interviews with patients was performed. Based on the secondary interviews, a second questionnaire was created for a matched case-control study. Through this analysis, prepackaged cookie dough was determined to be the source of the outbreak. Prepackaged cookie dough had previously never been implicated in a foodborne disease outbreak and therefore was not included in the first questionnaire (6). Similarly, in a 2011 nation-wide listeriosis outbreak, cantaloupe was determined to be the vehicle of transmission, even though produce is rarely contaminated by this pathogen (7). As these novel outbreaks become more common, targeting potential food sources based on pathogen may become less effective in outbreak investigations.

Several factors can affect the quantity and variety of food an individual or family consumes, including income, gender, age, and health status (8-16). Most of the research assessing food consumption patterns has occurred outside of the United States, and therefore not directly applicable (9, 10, 12). However, gender, age and state of residence based differences have been recently observed in food consumption using national survey data (8-16).

Men have been found more likely to consume 'high risk' foods, including unpasteurized milk, raw shellfish, runny eggs, and pink hamburgers when compared with women (8-10). In multiple studies it has been shown that men are more likely to eat meat when compared to women; moreover, this trend holds true even when looking at school-aged boys and girls (10-12). Women are more likely to consume fruits and vegetables, including alfalfa sprouts- also considered to be a 'high risk' food (10). Differences in consumption patterns between school-aged children and adults have also been shown in national surveys; children are more likely to consume raw fruits and tomatoes, and adults more likely to eat dark green vegetables, cantaloupe, and strawberries (13).

There is the potential to use consumption patterns to better predict and assist in the identification of contamination sources in foodborne disease outbreaks. In a 2004 *Salmonella enterica* (serotype Enteritidis) outbreak in Oregon, hypothesis-generating interviews proved unsuccessful in identifying a common exposure among patients (17). However, when the rate of food consumption of Oregon residents was compared to consumption estimates for cases, raw almonds was determined to be the source (17). In 2008, the source of an outbreak of *Salmonella enterica* serotype Agona infections was determined by comparing the percentage of cases reported eating puffed rice cereal with the total ready-to-eat cereal market share in the US (18).

There is a clear gap in knowledge concerning the foods associated with persons who have illness by demographics. Using data from foodborne disease outbreaks, this study analyzes the associations between food outbreak source, age and gender and the potential to use demographic data to assist in the identification the source of the outbreak.

Methods

Since 1973, the Centers for Disease Control and Prevention (CDC) has collected data on foodborne disease outbreaks (defined as 2 or more persons with a similar illness and exposure to a common food) through the Foodborne Disease Outbreak Surveillance System (FDOSS). FDOSS is a passive national surveillance system with state, local, territorial, and national public health agencies reporting outbreaks caused by bacterial, viral, parasitic, and chemical agents (2, 19).

Data collected for each outbreak includes the implicated food, number of illnesses, number and/or proportion of men and women, number and/or proportion of cases aged <4 years old, 5-19 years old, 20-49 years old, and over 50 years old, reporting state, and etiologic agent. We reviewed foodborne disease outbreaks reported to FDOSS where the first illness in the outbreak occurred during January 1, 1998-December 31, 2011.

For analysis, foods were further categorized into 17 mutually exclusive commodity groups (19). Proportions of men, women, and each age category were calculated using reported counts when proportions were not directly reported by the health department. Outbreak reports missing information on age or gender and those with inaccurate data (i.e., proportions added to <98% or >102%) were not included in the analysis. Reporting states were categorized into US Census divisions and regions. To adjust for outbreak size, a weighting scheme was created using quintile cutoff points (see Table 1).

Mean proportions for gender, age category, and US Census division were calculated for each commodity type. Additionally, crude and adjusted odds ratios were calculated for gender and age by commodity type. Age-adjusted odds ratios were calculated for gender by determining if the outbreak had more men or women , defined by a proportion of >50%; gender-adjusted odds ratios were calculated for age by determining if the outbreak had more than the expected mean proportion for each age group. All data analysis was performed using SAS 9.3 (SAS Institute., Cary, NC, USA).

Classification tree-based models for outbreak commodities were created utilizing gender, age categories, and region to determine if demographic information could help predict commodity sources of outbreaks. Tree models were weighted based on total case quintiles as described above (Table 1). Commodity groups were collapsed in numerous ways using the commodity hierarchy described by Painter et al (19).

Classification trees were created using RStudio 0.98.501 (RStudio, Inc.) using the RPART package (20). The minimum number of observations in a node before splitting (minsplit) was set to 30; cost complexity factor (cp) was set to .001. Trees were pruned by selecting a tree size, and associated cp, that minimized cross-validated error. Trees were created using both the complete dataset and datasets stratified by etiology. Kappa statistics, sensitivity, and specificity were calculated for each tree to assess potential predictability; the final model was selected based on these criteria. Institutional Review Board (IRB) exemption was received from Emory University.

Results

Between 1998 and 2011 there were 14,952 outbreaks reported to FDOSS. These outbreaks caused 303,662 total reported illnesses (median 8, range 2-1,939). Overall, a large portion of the data was missing or incomplete; 21.8% of gender data and 33.9% of age data were missing or misreported. Only 3,601 outbreaks (24.1%) identified a single commodity food source and were included in the analysis.

On average these outbreaks involved 44.2% males and 55.7% women (sd 26.1); 3.0% of cases were under the age of 4 (sd 10.5), 15.5% between 5 and 19 (sd 26.0), 52.7% between 20 and 49 (sd 33.2) and 26.8% over 50 years old (sd 31.4). See Table 2 for a complete breakdown of outbreaks, cases, age distributions and gender distributions by commodity type. Regionally, 1,151 (32.0%) of the included outbreaks were reported by states in the West, 735 (20.4%) were reported by states in the Midwest, 1,090 (30.3%) were reported by Southern states, and 463 (12.9%) were reported by Northeastern states (Table 3).

Unadjusted odds ratios were calculated for gender and each age group. Men were statistically significantly more likely to be involved in mollusk, game and vine-stalk outbreaks compared to other outbreak commodity types, odds ratio (OR) 1.51 (95% CI 1.36, 1.68), 1.83 (95% CI 1.29, 2.60), and 1.48 (95% CI 1.40, 1.57), respectively. Women were more likely to be involved in oil-sugar, fruit-nuts, leafy, and sprout outbreaks, OR 1.91 (95% CI 1.42, 2.57), 1.64 (95% CI 1.56, 1.73), 1.43 (95% CI 1.36, 1.51), and 1.53 (95% CI 1.29, 1.81), respectively.

Children under the age of 4 were more likely to be involved in dairy (OR=2.75, 95% CI 2.42, 3.12) and fruit-nuts (OR=4.55, 95% CI 4.14, 5.00) outbreaks; children between 5 and 19 were also more likely to be involved in dairy outbreaks (OR=1.99, 95% CI 1.83, 2.17), as well as game (OR=2.95, 95% CI 2.09, 4.15), poultry (OR= 1.51, 95% CI 1.42, 1.60),and grain-bean outbreaks (OR=1.63, 95% CI 1.48, 1.80). The odds of adults aged 20-49 being involved in a fish, crustacean, mollusk, sprout or vine-stalk vegetables outbreak were higher than other commodity types, ORs 2.05 (95% CI 1.88, 2.24), 1.78 (95% CI 1.48, 2.14), 1.81 (95% CI 1.61, 2.04), 2.07 (95% CI 1.76, 2.43), and 1.86 (95% CI 1.72, 2.01), respectively. Older adults, over 50 years old, were more likely to be involved in oil-sugar outbreaks (OR=3.21, 95% CI 2.38, 4.33), fruit-nuts outbreaks (OR=1.50, 95% CI 1.42, 1.59), and fungi outbreaks (OR=4.64, 95% CI 3.32, 6.48). See Table 4 for complete breakdown of odds ratios and confidence intervals for gender and age at each commodity level.

After adjusting for age, mollusk, dairy and meat (beef, game, and pork) outbreaks were more likely to have more men than women than outbreaks caused by other foods, aOR= 2.41 (95% CI 1.92, 3.03), 1.60 (95% CI 1.31, 1.95), 1.43 (95% CI 1.24, 1.64), 7.66 (95% CI 3.68, 15.94), and 1.46 (95% CI 1.23, 1.72), respectively. After adjustment, women were more like to be involved in fruit-nuts outbreaks (aOR 2.61, 95% CI 2.13, 3.20), leafy outbreaks (aOR 1.85, 95% CI 1.57, 2.18), sprout outbreaks (aOR 3.35, 95% CI 1.97, 5.70) and vine-stalk vegetables outbreaks (aOR 1.53, 95% CI 1.14, 2.05).

Children under 4 and 5-19 were more likely to be involved in outbreaks involving dairy, aOR=4.10 (95% CI 3.37, 5.00) and 3.47 (95% CI 2.85, 4.23), respectively, beef aOR= 1.50 (95% CI 1.20, 1.78) and 1.51 (95% CI 1.31, 1.73), respectively, pork, aOR= 1.46 (95% CI 1.42, 1.78) and 1.45 (95% CI 1.23, 1.71), respectively, and poultry, aOR= 1.30 (95% CI 1.13, 1.50) and 1.24 (95% CI 1.11, 1.40), respectively, when compared without breaks caused by other foods and after adjustment for gender. Additionally, children under 4 were more likely to be involved in egg, fruit-nuts and sprout outbreaks, aOR= 1.61 (95% CI 1.23, 2.09), 1.70 (95% CI 1.37, 2.10), and 4.54 (95% CI 2.88, 7.16), respectively, after adjustment. Children between 5 and 19 were also more likely to be involved in game outbreaks (aOR 2.12, 95% CI 1.19, 3.80) after adjustment for gender.

Adults (aged 20-49) were more likely to be involved in outbreaks caused by aquatic animals- fish, crustaceans, and mollusks- aOR 1.61 (95% CI 1.40, 1.85), 1.78 (95% CI 1.24, 2.58), and 1.57 (95% CI 1.25, 1.98), respectively, as well as grain-bean (aOR 1.47, 95% CI 1.18, 1.83), leafy (aOR 1.38, 95% CI 1.19, 1.61), and sprout (aOR 4.99, 95% CI 2.74, 9.06) outbreaks after gender adjustment. Older adults, (> 50), were also more likely to be involved in fish outbreaks (aOR 1.54, 95% CI 1.29, 1.83), as well as oil-sugar (aOR 2.78, 95% CI 1.24, 6.23), fruit-nuts (aOR 1.78, 95% CI 1.45, 2.20), root vegetable (aOR 2.18, 95% CI 1.45, 3.28) and vine-stalk vegetables outbreaks (aOR 1.68, 95% CI

1.20, 2.36). See Table 5 for a complete breakdown of adjusted odds ratios for gender and age by commodity type.

The final classification tree initially contained 52 nodes, but was pruned to include only 13 nodes for ease of use (Figure 1). Commodities were collapsed from seventeen categories to six in the final model: aquatic (fish, crustaceans and mollusks), dairy/eggs, meat (beef, game, and pork), poultry, plants (grain-bean and oil-sugar), and produce (fruit-nuts, fungi, leafy, root, sprout, and vine-stalk vegetables). A Kappa agreement statistic of .17 was achieved for the pruned tree; the full tree had a .23 agreement. Sensitivity and specificity was calculated for each commodity, sensitivities ranged from 0%-57.4% and specificities 78.1%-100%. See Table 6 for breakdown of sensitivity and specificity results for each commodity. Splits were created based off the weighted proportions of each category in the data. For example, the first split separates outbreaks where cases between 20 and 49 years old make up at least 96% of the outbreak and outbreaks where this age group makes up less than 96% of cases (Figure 1).

Discussion

Outbreaks caused by meat were more likely to involve more men than women, and outbreaks caused by fruit-nuts and vegetables (specifically, leafy vegetables, sprouts, and vine-stalk vegetables) were more likely to involve more women than men. The age distribution in outbreaks also varied by food; more children were affected in outbreaks caused by dairy and poultry while more adults were involved in outbreaks caused by vegetables. Current literature on food consumption patterns in the United States support that certain age groups or gender are more likely to be associated with specific food types (8-16). While the classification tree model provided minimal predictability, this preliminary analysis provides evidence that there is potential to use demographic data to predict contamination source.

These demographic associations could potentially assist outbreak investigations by providing information about foods that are more likely to be associated with certain demographic groups. In the two recent milk outbreaks, over 30% of cases in each were in children, twice the average proportion of children in foodborne disease outbreaks; these outbreaks both fit the current findings that children are more likely to be involved in outbreaks with contaminated dairy (21, 22). In the listeriosis outbreak caused by cantaloupes described previously, 86% of patients were over the age of 60 and 58% of cases were female, both findings which fit the results of the current analysis (7).

To our knowledge, this is the first commodity classification tree created to help predict outbreak source contamination. A tree model using only demographic information (age, gender, and region) can provide utility as this information is typically readily available to investigators during an outbreak investigation. However, the current model may not be suitable for use in its current form due to its limited predictability. Additional demographic information and a more comprehensive dataset may help to create a better fitting predictive model. Moreover, a predictive classification tree model will need to be validated in outbreak investigations as a viable tool before it should be used in practice.

Limitations to these findings are due to inherent constraints of the surveillance data used. First, gender and age are reported in aggregate and therefore interactions between the two variables could not be assessed. Future studies could refine these estimates by using datasets where age and gender of each case is known so a more accurate assessment can be made of gender and age on outbreak source. Second, as a large percentage of the data is missing, incomplete, or misreported, it is possible that the missing data could limit generalizability and predictability of the tree model.

Third, foodborne disease outbreaks represent only a small portion of total foodborne illnesses; moreover, only a minority of illnesses are reported to health departments, and therefore FDOSS. These findings and associations may not apply to all foodborne illnesses, and rather represent only the small subset of foodborne disease outbreaks that are reported. However, this study uses 14 years of data from a large nationwide outbreak surveillance dataset with a standardized food classification scheme to allow for meaningful analysis. This study, to the best of the authors' knowledge, is the first to use a dataset of this nature to identify demographic groups associated with foods causing outbreaks.

Knowledge on associations between outbreak demographic characteristics and contaminated food source is still limited. As novel contamination routes and outbreak sources continue to become more prevalent, new methods for outbreak source identification will need to be developed. This study provides preliminary evidence that demographic information may be able to assist investigators in identifying contaminated food sources in outbreaks. Additional research can provide a more concrete sense of the predictive power of demographic data in these investigations.

References

1. Scallan E, Hoekstra RM, Angulo FJ, Tauxe RV, Widdowson MA, Roy SL, Jones JL, Griffin PM. Foodborne illnesses acquired in the United States- major pathogens. *Emerg Infect Dis*. 2011; 17(1): 7-15.

2. Gould LH, Walsh KA, Vieira AR, Herman K, Williams IT, Hall AJ and Cole D. Surveillance for foodborne disease outbreaks—United States, 1998-2008. *MMWR Surveill Summ*. 2013; 62(2):1-34.

3. Centers for Disease Control and Prevention (CDC). Surveillance for foodborne disease outbreaks—United States, 2009-2010. *MMWR Morb Mortal Wkly Rep*. 2013; 62(3):41-47.

4. Centers for Disease Control and Prevention. Generating hypotheses about likely sources. http://www.cdc.gov/foodsafety/outbreaks/investigating-outbreaks/investigations/sources.html. Updated November 12, 2013. Accessed December 2, 2013.

5. Pires SM, Evers EG, van Pelt W, Ayers T, Scallan E, Angulo FJ, Havelaar, Hald T, Med-Vet-Net Workpackage 28 Working Group. Attributing the human disease burden of foodborne infections to specific sources. *Foodborne Pathog Dis*. 2009; 6(4): 417-424.

6. Neil KP, Biggerstaff G, MacDonald JK, Trees E, Medus C, Musser KA, Stroika SG, Zink D and Sotir MJ. A novel vehicle for transmission of *Escherichia coli* O157:H7 to humans: Multistate outbreak of *E. coli* O157:H7 infections associated with consumption of ready-to-break commercial prepackaged cookie dough-United States, 2009. *Clin Infect Dis*. 2012; 54(4): 511-518.

7. McCollum JT, Cronquist AB, Silk BJ, et al. Multistate outbreak of Listeriosis associated with cantaloupe. *N Engl J Med*. 2013; 369(10): 944-953.

8. Shiferaw B, Yang S, Cieslak P, Vugia D, Marcus R, Koehler J, Deneen V, Angulo F, the FoodNet Working Group. Prevalence of high-risk food consumption and food-handling practices among adults: A multi-states survey, 1996 to 1997. *J Food Protection*. 2000; 63(11): 1538-1543.

9. Samuel MC, Vugia DJ, Koehler KM, Marcus R, Deneen V, Damaske B, Shiferaw B, Hadler J, Henao OL, Angulo FJ. Consumption of risky foods among adults at high risk for severe foodborne diseases: Room for improved targeted prevention messages. *J Food Safety*. 2007; 27: 219-232.

10. Shiferaw B, Verrill L, Booth H, Zansky SM, Norton DM, Crim S, Henao OL. Sex-based differences in food consumption: Foodborne diseases active surveillance network (FoodNet) population survey, 2006-2007. *Clin Infect Dis*. 2012; 54 Suppl 5: S453-S457.

11. Daniel CR, Cross AJ, Koebnick C, Sinha R. Trends in meat consumption in the USA. *Public Health Nutr*. 2011; 14(4): 575-583.

12. Wardle J, Haase AM, Steptoe A, Nillapun M, Jonwutiwes K, Bellisle F. Gender differences in food choice: the contribution of health beliefs and dieting. *Ann Behav Med*. 2004; 27(2): 107-116.

13. Caine-Bish NL, Scheule B. Gender differences in food preferences of school-aged children and adolescents. *J Sch Health*. 2009; 79(11): 532-540.

14. Kimmons J, Gillespie C, Seymour J, Serdula M, Blanck HM. Fruit and vegetable intake among adolescents and adults in the United States: Percentage meeting individualized recommendations. *Medscape J Med*. 2009; 11(1): 26.

15. Emanuel AS, McCully SN, Gallagher KM, Updegraff JA. Theory of planned behavior explains gender difference in fruit and vegetable consumption. *Appetite*. 2012; 59 693-697.

16. Dehghan M, Akhtar-Danesh N, Merchant AT. Factors associated with fruit and vegetable consumption among adults. *J Hum Nutr Diet*. 2011; 24(2): 128-134.

17. Centers for Disease Control and Prevention (CDC). Outbreak of Salmonella serotype Enteritidis infections associated with raw almonds—United States and Canada, 2003-2004. *MMWR Morb Mortal Wkly Rep*. 2004; 53(22): 484-487.

18. Russo ET, Biggerstaff G, Hoekstra RM, Meyer S, Patel N, Miller B, Quick R. A recurrent, multistate outbreak of *Salmonella* serotype Agona infections associated with dry, unsweetened cereal consumption, United States, 2008. *J Food Protect*. 2013; 76(2): 227-230.

19. Painter JA, Ayers T, Woodruff R, Blanton E, Perez N, Hoekstra RM, Griffin PM and Braden C. Recipes for foodborne outbreaks: A scheme for categorizing and grouping implicated foods. *Foodborne Pathog Dis*. 2009; 6(10): 1259-1264.

20. Therneau TM, Atkinson EJ. An introduction to recursive partitioning using the RPART routines. Technical Report 61, Section of Biostatistics, Mayo Clinic, Rochester. 1997. Available at http://cran.r-project.org/web/packages/rpart/vignettes/longintro.pdf

21. Longenberger AH, Palumbo AJ, Chu AK, Moll ME, Weltman A, Ostroff SM. *Campylobacter jejuni* infections associated with unpasteurized milk- multiple states, 2012. *Clin Infect Dis*. 2013; 57(2): 263-266.

22. Centers for Disease Control and Prevention (CDC). Recurrent outbreak of *Campylobacter jejuni* infections associated with raw milk dairy- Pennsylvania, April-May 2013. *MMWR Morb Mortal Wkly Rep*. 2013; 62(34):702.

Table 1. Outbreak Weighting

Outbreak Size	Weight
≤2 cases	1
2-4 cases	2
4-10 cases	3
10-25 cases	4
>25 cases	5

Table 2. Age and gender case distributions by food commodity, Foodborne Disease Outbreak Surveillance System, 1998-2011.

Commodity	Outbreaks	Illnesses	Male Cases (%)[a]	Female Cases (%)	Cases ≤4 years (%)	Cases 5-19 years (%)	Cases 20-49 years (%)	Cases 50+ years (%)
Fish	649	3,560	1,201 (48)	1,438 (52)	35 (1)	133 (6)	1,568 (50)	629 (26)
Crustaceans	93	912	290 (46)	324 (54)	6 (1)	47 (10)	314 (66)	126 (27)
Mollusks	212	2,700	1,046 (60)	690 (40)	6 (0)	59 (5)	782 (64)	375 (26)
Dairy	198	5,815	1,840 (54)	1,551 (46)	293 (10)	812 (29)	1,205 (64)	472 (31)
Eggs	153	5,541	1,316 (49)	1,385 (51)	174 (8)	420 (19)	963 (43)	590 (17)
Beef	437	9,220	3,490 (50)	3,441 (50)	145 (3)	1,167 (23)	2,324 (45)	1,364 (27)
Game	26	224	121 (65)	66 (35)	3 (2)	54 (39)	61 (44)	19 (14)
Pork	291	5,812	2,265 (50)	2,305 (50)	114 (3)	642 (18)	1,774 (49)	1,090 (30)
Poultry	658	12,525	5,355 (51)	5,165 (49)	207 (3)	1,742 (23)	3,503 (47)	2,039 (27)
Grain-Bean	183	3,291	1,611 (57)	1,223 (43)	25 (1)	568 (26)	1,233 (56)	382 (17)
Oil-Sugar	12	246	80 (35)	147 (64)	2 (1)	33 (19)	43 (25)	95 (55)
Fruit-Nuts	211	9,319	2,887 (39)	4,560 (61)	768 (14)	816 (15)	2,038 (36)	1,966 (35)
Fungi	22	130	46 (48)	49 (52)	0 (0)	11 (14)	43 (54)	26 (33)
Leafy	284	9,513	2,921 (41)	4,196 (59)	79 (1)	701 (13)	2,981 (54)	1,704 (31)
Root	49	1,293	639 (55)	517 (45)	15 (2)	162 (18)	495 (54)	250 (27)
Sprout	39	1,327	414 (45)	502 (55)	23 (3)	106 (16)	445 (66)	90 (13)
Vine-stalk vegetables	84	7,433	3,232 (57)	2,408 (43)	63 (2)	279 (9)	1,925 (64)	740 (25)
Total[b]	3,601	78, 861	28,754 (49)	29,967 (51)	1,958 (5)	7,752 (18)	21, 697 (50)	11,957 (27)

[a] Percentage of all cases due to that commodity in the indicated group, calculated using non-missing data only
[b] Total includes all outbreaks where a single food commodity was identified

Table 3. Region and division outbreak distributions (and percentage)[a], by food commodity, Foodborne Disease Outbreak Surveillance System between 1998 and 2011.

Commodity	Total (N=13,648)	West		Midwest			South		Northeast	
		Pacific[b]	Mountain	West South Central	East South Central	South Atlantic	West North Central	East North Central	Mid-Atlantic	New England
Fish	649	309 (48)	13 (2)	18 (3)	30 (5)	3 (0)	4 (1)	177 (28)	70 (11)	16 (3)
Crustaceans	93	11 (12)	2 (2)	3 (3)	5 (5)	1 (1)	1 (1)	61 (67)	6 (7)	1 (1)
Mollusks	212	85 (41)	5 (2)	5 (2)	2 (1)	13 (6)	5 (2)	70 (34)	16 (8)	6 (3)
Dairy	198	42 (22)	27 (14)	21 (11)	26 (14)	8 (4)	3 (2)	29 (15)	19 (10)	13 (7)
Eggs	153	30 (22)	22 (8)	10 (7)	22 (16)	1 (1)	2 (1)	30 (22)	20 (15)	10 (7)
Beef	437	72 (18)	22 (5)	35 (9)	75 (19)	10 (2)	13 (3)	111 (28)	55 (14)	10 (2)
Game	26	11 (44)	1 (4)	1 (4)	3 (12)	0 (0)	2 (8)	2 (8)	4 (16)	1 (4)
Pork	291	46 (16)	10 (4)	21 (7)	51 (18)	3 (1)	13 (5)	112 (39)	19 (7)	10 (4)
Poultry	658	149 (23)	17 (3)	42 (7)	103 (16)	21 (3)	17 (3)	223 (35)	53 (8)	19 (3)
Grain-Bean	183	55 (31)	13 (7)	11 (6)	15 (8)	1 (1)	5 (3)	63 (35)	17 (9)	0 (0)
Oil-Sugar	12	3 (25)	0 (0)	2 (17)	1 (8)	0 (0)	0 (0)	6 (50)	0 (0)	0 (0)
Fruit-Nuts	211	40 (21)	12 (6)	29 (15)	31 (16)	3 (2)	3 (2)	49 (26)	14 (7)	9 (5)
Fungi	22	7 (32)	3 (14)	1 (5)	2 (9)	0 (0)	0 (0)	3 (14)	5 (23)	1 (5)
Leafy	284	72 (27)	21 (8)	20 (8)	49 (19)	4 (2)	10 (4)	44 (17)	26 (10)	17 (6)
Root	49	14 (29)	2 (4)	4 (8)	4 (8)	1 (2)	0 (0)	16 (33)	4 (8)	3 (6)
Sprout	39	8 (36)	3 (14)	2 (9)	3 (14)	0 (0)	0 (0)	4 (18)	0 (0)	2 (9)
Vine-stalk vegetables	84	17 (25)	11 (16)	3 (4)	11 (16)	4 (6)	3 (4)	7 (10)	6 (9)	7 (10)
Total[c]	3,601	976 (28)	175 (5)	229 (7)	433 (13)	73 (2)	81 (2)	1,009 (29)	337 (10)	126 (4)

[a] Percentage of all outbreaks associated with the listed commodity in the indicated region, calculated using non-missing data only

[b] Pacific: Alaska, California, Hawaii, Oregon and Washington; Mountain: Arizona, Colorado, Idaho, Montana, Nevada, New Mexico, Utah, Wyoming; West South Central: Arkansas, Louisiana, Oklahoma, Texas; East South Central: Alabama, Kentucky, Mississippi, Tennessee; South Atlantic: Delaware, Georgia, Florida, Maryland, North Carolina, South Carolina, Virginia, West Virginia, Washington D.C.; West North Central: Iowa, Kansas, Minnesota, Montana, Nebraska, North Dakota, South Dakota; East North Central: Ohio, Illinois, Indiana, Michigan, Ohio; Mid-Atlantic: New Jersey, New York, Pennsylvania; New England: Connecticut, Maine, Massachusetts, New Hampshire, Rhode Island, Vermont

[c] Total includes all outbreaks where a single food commodity was identified

Table 4. Crude odds ratios (ORs) for gender and age categories by food commodity, Foodborne Disease Outbreak Surveillance System, 1998-2011.

Commodity	OR Male (95% CI)[a]	OR Female (95% CI)	OR ≤4 (95% CI)	OR 5-19 (95% CI)	OR 20-49 (95% CI)	OR > 50 (95% CI)
Fish	0.87 (0.80,0.94)	**1.15 (1.06, 1.25)**	0.31 (0.23, 0.44)	0.26 (0.22, 0.31)	**2.05 (1.88, 2.24)**	0.94 (0.86, 1.04)
Crustaceans	0.90 (0.76, 1.06)	1.11 (0.95, 1.31)	0.25 (0.11, 0.55)	0.48 (0.36, 0.65)	**1.78 (1.48, 2.14)**	0.90 (0.73, 1.10)
Mollusks	**1.51 (1.36, 1.68)**	0.66 (0.60, 0.73)	0.11 (0.05, 0.24)	0.29 (0.18, 0.30)	**1.81 (1.61, 2.04)**	**1.16 (1.03, 1.32)**
Dairy	**1.28 (1.19, 1.38)**	0.78 (0.73, 0.84)	**2.75 (2.42, 3.12)**	**1.99 (1.83, 2.17)**	0.75 (0.69, 0.81)	0.51 (0.46, 0.57)
Eggs	1.00 (0.92, 1.08)	1.00 (0.92, 1.09)	**1.89 (1.61, 2.22)**	**1.13 (1.01, 1.26)**	0.81 (0.74, 0.88)	0.99 (0.90, 1.09)
Beef	**1.09 (1.03, 1.15)**	0.92, (0.87, 0.97)	0.65 (0.55, 0.77)	**1.47 (1.37, 1.58)**	0.85 (0.80, 0.90)	0.98 (0.91, 1.04)
Game	**1.83 (1.29, 2.60)**	0.55 (0.38, 0.78)	0.75 (0.31, 1.83)	**2.95 (2.09, 4.15)**	0.79 (0.56, 1.10)	0.41 (0.26, 0.67)
Pork	1.00 (0.94, 1.07)	1.00 (0.94, 1.07)	0.72 (0.60, 0.87)	0.99 (0.91, 1.08)	0.96 (0.89, 1.02)	**1.13 (1.05, 1.22)**
Poultry	**1.16 (1.11, 1.21)**	0.86 (0.82, 0.90)	0.59 (0.51, 0.68)	**1.51 (1.42, 1.60)**	0.86 (0.81, 0.90)	0.97 (0.92, 1.02)
Grain-Bean	**1.29 (1.19, 1.39)**	0.78 (0.72, 0.84)	0.33 (0.24, 0.46)	**1.63 (1.48, 1.80)**	**1.27 (1.17, 1.39)**	0.53 (0.47, 0.59)
Oil-Sugar	0.52 (0.39, 0.70)	**1.91 (1.42, 2.57)**	0.23 (0.06, 0.95)	1.09 (0.74, 1.59)	0.33 (0.23, 0.47)	**3.21 (2.38, 4.33)**
Fruit-Nuts	0.61 (0.58, 0.64)	**1.64 (1.56, 1.73)**	**4.55 (4.14, 5.00)**	0.76 (0.71, 0.83)	0.53 (0.50, 0.57)	**1.50 (1.42, 1.59)**
Fungi	0.93 (0.59, 1.48)	1.07 (0.68, 1.71)	NA[b]	0.37 (0.20, 0.68)	0.41 (1.17, 1.31)	**4.64 (3.32, 6.48)**
Leafy	0.70 (0.66, 0.73)	**1.43 (1.36, 1.51)**	0.27 (0.22, 0.34)	0.65 (0.60, 0.70)	**1.24 (1.17, 1.31)**	**1.22 (1.14, 1.29)**
Root	**1.40 (1.23, 1.58)**	0.72 (0.63, 0.81)	0.35 (0.21, 0.57)	0.98 (0.83, 1.17)	**1.17 (1.03, 1.33)**	0.97 (0.84, 1.13)
Sprout	0.66 (0.55, 0.78)	**1.53 (1.29, 1.81)**	0.72 (0.47, 1.09)	0.88 (0.71, 1.08)	**2.07 (1.76, 2.43)**	0.41 (0.33, 0.51)
Vine-stalk vegetables	**1.48 (1.40, 1.57)**	0.67 (0.64, 0.71)	0.44 (0.34, 0.56)	0.45 (0.40, 0.51)	**1.86 (1.72, 2.01)**	0.84 (0.78, 0.92)

Bold lettering indicates an OR > 1 and significant at a 0.05 level
[a] 95% CI = 95% confidence interval
[b] No fungi outbreaks were observed in children under 4, therefore no odds ratio was calculated

Table 5. Weighted adjusted odds ratios (aOR) for gender and age categories by food commodity, Foodborne Disease Outbreak Surveillance System, 1998-2011.

Commodity	aOR Male[a] (95% CI)[c]	aOR Female (95% CI)	aOR ≤4[b] (95% CI)	aOR 5-19 (95% CI)	aOR 20-49 (95% CI)	aOR >50 (95% CI)
Fish	1.06 (0.91, 1.23)	0.53 (0.46, 0.61)	0.28 (0.21, 0.38)	0.33 (0.28, 0.40)	**1.61 (1.40, 1.85)**	**1.54 (1.29, 1.83)**
Crustaceans	0.61 (0.41, 0.93)	0.86 (0.61, 1.23)	0.69 (0.40, 1.19)	0.87 (0.60, 1.28)	**1.78 (1.24, 2.58)**	1.01 (0.62, 1.65)
Mollusks	**2.41 (1.92, 3.03)**	0.29 (0.22, 0.37)	0.30 (0.19, 0.49)	0.34 (0.25, 0.47)	**1.57 (1.25, 1.98)**	**1.62 (1.23, 2.15)**
Dairy	**1.60 (1.31, 1.95)**	0.58 (0.47, 0.71)	**4.10 (3.37, 5.00)**	**3.47 (2.85, 4.23)**	0.41 (0.34, 0.51)	0.50 (0.35, 0.71)
Eggs	0.87 (0.68, 1.11)	0.99 (0.79, 1.25)	**1.61 (1.23, 2.09)**	1.04 (0.82, 1.31)	0.65 (0.51, 0.81)	1.30 (0.97, 1.74)
Beef	**1.43 (1.24, 1.64)**	0.80 (0.69, 0.91)	**1.50 (1.27, 1.78)**	**1.51 (1.31, 1.73)**	0.72 (0.62, 0.82)	0.95 (0.78, 1.18)
Game	**7.66 (3.68, 15.94)**	0.23 (0.11, 0.47)	1.18 (0.52, 2.68)	**2.12 (1.19, 3.80)**	0.62 (0.35, 1.13)	0.83 (0.33, 2.11)
Pork	**1.46 (1.23, 1.72)**	0.80 (0.68, 0.94)	**1.46 (1.20, 1.78)**	**1.45 (1.23, 1.71)**	0.93 (0.79, 1.09)	0.84 (0.66, 1.07)
Poultry	0.92 (0.81, 1.04)	1.09 (0.97, 1.22)	**1.30 (1.13, 1.50)**	**1.24 (1.11, 1.40)**	0.91 (0.82, 1.03)	0.78 (0.65, 0.92)
Grain-Bean	1.00 (0.80, 1.25)	1.06 (0.85, 1.32)	0.50 (0.35, 0.73)	1.06 (0.84, 1.33)	**1.47 (1.18, 1.83)**	0.52 (0.35, 0.76)
Oil-Sugars	1.73 (0.80, 3.74)	1.08 (0.50, 2.31)	1.16 (0.44, 3.06)	2.12 (0.99, 4.52)	0.35 (0.15, 0.83)	**2.78 (1.24, 6.23)**
Fruit-Nuts	0.41 (0.32, 0.52)	**2.61 (2.13, 3.20)**	**1.70 (1.37, 2.10)**	1.03 (0.85, 1.24)	0.59 (0.49, 0.71)	**1.78 (1.45, 2.20)**
Fungi	1.33 (0.67, 2.63)	0.41 (0.20, 0.85)	NA[d]	1.41 (0.72, 2.76)	0.65 (0.33, 1.28)	1.24 (0.51, 3.01)
Leafy	0.65 (0.55, 0.78)	**1.85 (1.57, 2.18)**	0.48 (0.37, 0.63)	0.68 (0.57, 0.81)	**1.38 (1.19, 1.61)**	1.03 (0.84, 1.27)
Root	1.23 (0.84, 1.79)	1.35 (0.93, 1.95)	0.67 (0.39, 1.18)	1.09 (0.75, 1.59)	0.70 (0.48, 1.01)	**2.18 (1.45, 3.28)**
Sprout	0.40 (0.23, 0.72)	**3.35 (1.97, 5.70)**	**4.54 (2.88, 7.16)**	1.32 (0.83, 2.10)	**4.99 (2.74, 9.06)**	NA[e]
Vine-stalk vegetables	0.85 (0.62, 1.15)	**1.53 (1.14, 2.05)**	1.14 (0.80, 1.65)	0.61 (0.44, 0.84)	1.06 (0.80, 1.41)	**1.68 (1.20, 2.36)**

Bold lettering indicates an OR > 1 and significant at a 0.05 level

[a] Gender variables (defined as >50% or ≤50%) were adjusted by age category proportions; age 20-49 was used as the referent group

[b] Age variables (defined as >mean proportion or ≤ mean proportion) were adjusted by gender proportions; female was used as the referent group

[c] 95% CI = 95% confidence interval

[d] No fungi outbreaks were observed in children under 4, therefore no adjusted odds ratio was calculated

[e] No sprout outbreaks had over 26.8% adults over 50, so no adjusted odds ratio was calculated

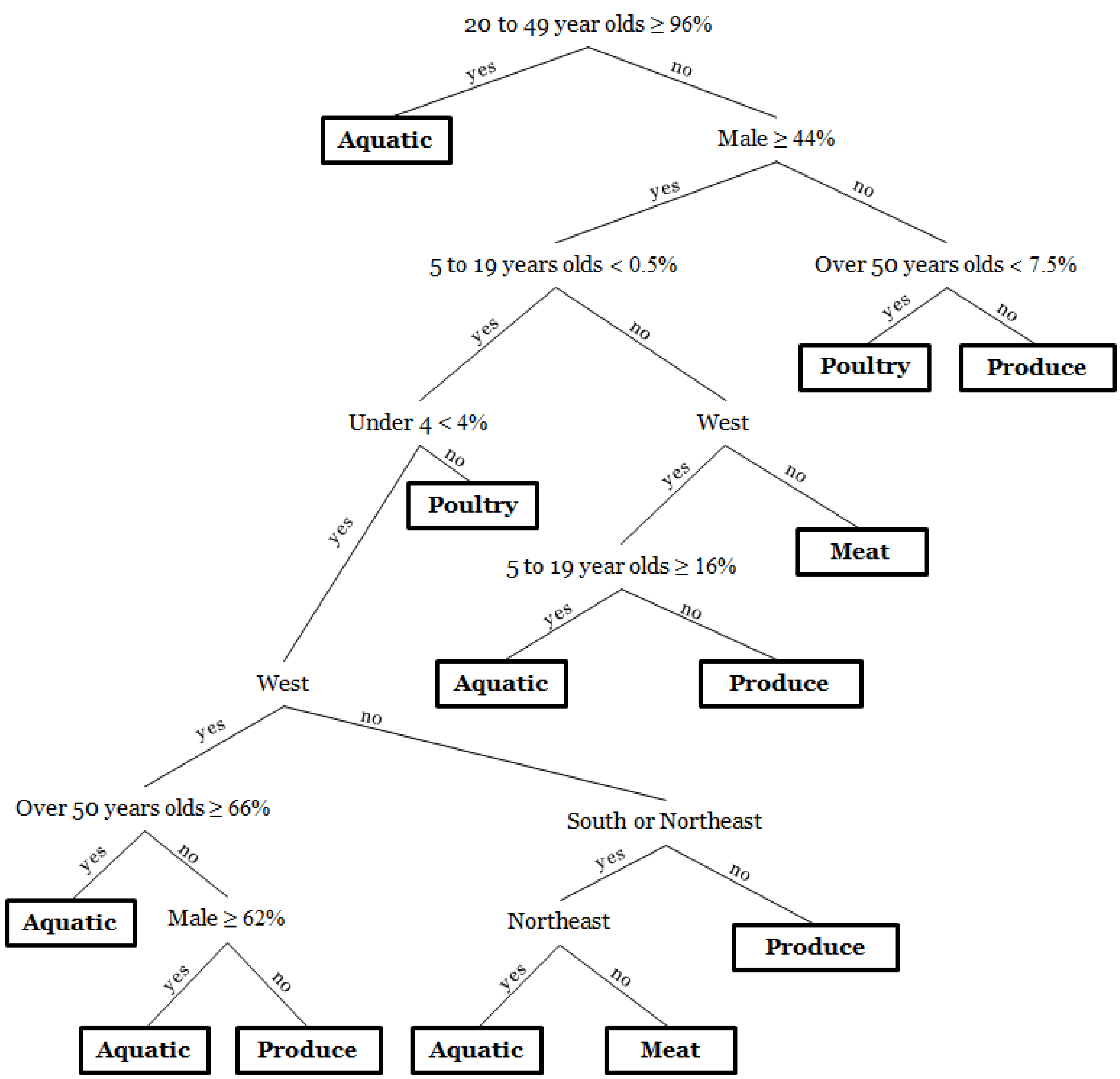
Table 6. Sensitivity and specificity for final classification tree-based model (Figure 1), by collapsed commodity type.

Commodity[a]	Aquatic	Dairy/Eggs[b]	Meat	Poultry	Plants[b]	Produce
Sensitivity (%)	57.4	0.0	42.2	12.8	0.0	47.2
Specificity (%)	80.0	100	78.1	93.5	100	81.5

[a] Commodity groups were collapsed to maximize Kappa; aquatic= fish, crustacean and mollusk; meat = beef, game and pork; plants = grain-bean, and oil-sugar; produce = fruit-nuts, fungi, leafy, root, sprout and vine-stalk vegetables

[b]Dairy/egg and plant outbreaks were not predicted in the final model

Figure 1. Classification decision tree to predict collapsed commodity type using gender, age and region.



Summary, Public Health Implications, Possible Future Directions

The average reported outbreak in the United States between 1998 and 2011 involved 44% men, 56% women, almost 19% children, 53% adults and 27% older adults (50+ years old). However, when data is stratified by contaminated food source, the average number of men, women and age groups vary from these overall estimates. Males were more likely to be the majority in outbreaks caused by meat, and females more likely to be involved in outbreaks caused by fruits and vegetables (specifically leafy, sprouts, and vine-stalk vegetables). Age was also associated with contaminated commodity; children were more likely to be involved in outbreaks caused by dairy and poultry and adults more likely to be involved in outbreaks caused by vegetables. While the classification tree model provided minimal predictability, kappa agreement=0.17, this preliminary analysis provides evidence that there is potential to use demographic data to predict contamination source.

Overall, this analysis of food outbreak data shows that there is an association between demographic features and contaminated food source. While etiology is strongly correlated with outbreak commodity (Appendix Table 6), outbreaks caused by previously unidentified food vehicles are making this association less useful in source identification. Additionally, an etiology is not always determined. By using information that is always available- basic case demographics- this could add another resource to the current tool kit used by outbreak investigators and health departments. If outbreak sources can be identified more rapidly and with greater frequency, the benefits would be two-fold: recalls could occur faster which would prevent additional exposure and illnesses, as well as assist in the furthering of current knowledge about foodborne disease transmission.

This is the first commodity classification tree created to help predict outbreak source contamination. A tree model using only demographic information (age, gender, and region) can provide utility as this information is readily available to investigators at the start of an outbreak investigation. However, the current model has limited predictability and may not be suitable for use in its current form. Additional demographic information and a more comprehensive dataset may help to create a better fitting predictive model. Moreover, a predictive classification tree model will need to be validated in outbreak investigations as a viable tool before it should be used in practice.

Limitations to these findings are due to inherent constraints of the surveillance data used. First, gender and age are reported in aggregate and therefore interactions between the two variables could not be assessed. Future studies could refine these estimates by using datasets where age and gender of each case is known so a more accurate assessment can be made of gender and age on outbreak source. Second, as a large percentage of the data is missing, incomplete, or misreported, it is possible that the missing data could limit generalizability and predictability of the tree model (Appendix Table 7).

Third, foodborne disease outbreaks represent only a small portion of total foodborne illnesses; moreover, only a minority of illnesses are reported to health departments, and therefore FDOSS. These findings and associations may not apply to all foodborne illnesses, and rather represent only the small subset of foodborne disease outbreaks that are reported. However, this study uses 14 years of data from a large nationwide outbreak surveillance dataset with a standardized food classification scheme to allow for meaningful analysis. This study, to the best of the authors' knowledge, is the first to use a dataset of this nature to identify demographic groups associated with foods causing outbreaks. Knowledge on associations between outbreak demographic characteristics and contaminated food source is still limited. As novel contamination routes and outbreak sources continue to become more prevalent, new methods for outbreak source identification will need to be developed. This study provides preliminary evidence that demographic information may be able to assist investigators in identifying contaminated food sources in outbreaks. Additional research can provide a more concrete sense of the predictive power of demographic data in these investigations.

Appendix

Figure 2. Institutional Review Board (IRB) exemption letter.

EMORY UNIVERSITY

Institutional Review Board

October 15, 2013

Paul Strassle
Rollins School of Public Health
Emory University
Atlanta, GA 30322

RE: Determination: No IRB Review Required
Title: Using Gender and Age Distributions in Foodborne Outbreak Cases to Predict Source of Contamination
eIRB#: 65819

Dear Mr. Strassle:

Thank you for requesting a determination from our office about the above-referenced project. Based on our review of the materials you provided, we have determined that it does not require IRB review because it does not meet the definition(s) of "research" involving "human subjects" as set forth in Emory policies and procedures and federal rules, if applicable. Specifically, in this project, the data used is available to the public upon request from the CDC. This data is collected in a self-reporting passive surveillance system utilized by local and state health departments, as well as the CDC, where outbreak information is uploaded. There is no active collection of data (surveys, questionnaires, etc.) or intervention.

Each entry in the database is an outbreak, so there are no personal identifiers or health information included. Moreover, due to the nature of the data there is no individual information included at all each row represents an outbreak (with a minimum of 2 cases) and all demographic information reported in proportions to the total outbreak population.

Please note that this determination does not mean that you cannot publish the results. If you have questions about this issue, please contact me.

This determination could be affected by substantive changes in the study design, subject populations, or identifiability of data. If the project changes in any substantive way, please contact our office for clarification.

Thank you for consulting the IRB.

Sincerely,

Brandy Covington
Research Protocol Analyst, Sr.
Emory University Institutional Review Board

Emory University
1599 Clifton Road, 5th Floor - Atlanta, Georgia 30322
Tel: 404.712.0720 - Fax: 404.727.1358 - Email: irb@emory.edu - Web: http://www.irb.emory.edu
An equal opportunity, affirmative action university

Table 6. Etiologic agents causing of outbreaks, by food commodity, Foodborne Disease Outbreak Surveillance System, 1998-2011.

Commodity	Outbreaks	Bacterial (%)[a]	Viral (%)	Chemical (%)	Parasitic (%)
Fish	649	44 (7)	8 (1)	535 (82)	1 (0)
Crustaceans	93	32 (34)	5 (5)	4 (4)	1 (2)
Mollusks	212	74 (35)	70 (33)	12 (6)	1 (0)
Dairy	198	155 (78)	19 (10)	0 (0)	0 (0)
Eggs	153	137 (90)	3 (2)	0 (0)	0 (0)
Beef	437	292 (67)	28 (6)	1 (0)	0 (0)
Game	26	16 (62)	1 (4)	1 (4)	7 (27)
Pork	291	183 (63)	26 (9)	1 (0)	1 (0)
Poultry	658	370 (56)	43 (7)	3 (0)	0 (0)
Grain-Bean	183	96 (52)	24 (13)	3 (2)	1 (1)
Oil-Sugar	12	2 (17)	5 (42)	1 (8)	0 (0)
Fruit-Nuts	211	61 (29)	89 (42)	12 (6)	9 (4)
Fungi	22	2 (9)	2 (9)	15 (68)	0 (0)
Leafy	284	60 (21)	163 (57)	2 (1)	4 (1)
Root	49	22 (45)	13 (27)	1 (2)	0 (0)
Sprout	39	37 (95)	0 (0)	0 (0)	0 (0)
Vine-stalk vegetables	84	48 (57)	18 (21)	1 (1)	0 (0)
Total[b]	3,601	1,631 (45)	517 (14)	592 (16)	25 (1)

[a] Percentage calculated using non-missing data only
[b] Total includes all outbreaks where a single food commodity was identified

Table 7. Number of outbreaks with missing, incomplete, or misreported[a] demographic data by commodity type, Foodborne Disease Outbreak Surveillance System, 1998-2011.

Commodity	Outbreaks	Outbreaks without Gender Data (%)	Outbreaks without Age Data (%)
Fish	649	126 (19)	164 (25)
Crustaceans	93	22 (24)	35 (38)
Mollusks	212	43 (20)	69 (33)
Dairy	198	28 (14)	47 (24)
Eggs	153	36 (24)	49 (32)
Beef	437	87 (20)	147 (34)
Game	26	4 (15)	9 (35)
Pork	291	61 (21)	94 (32)
Poultry	658	119 (18)	182 (28)
Grain-Bean	183	32 (17)	53 (29)
Oil-Sugar	12	3 (25)	4 (33)
Fruit-Nuts	211	44 (21)	72 (34)
Fungi	22	5 (23)	7 (32)
Leafy	284	52 (18)	88 (31)
Root	49	9 (18)	14 (29)
Sprout	39	11 (28)	15 (38)
Vine-stalk vegetables	84	15 (18)	30 (36)
Total[b]	3,601	687 (19)	1,079 (30)

[a] Data was considered misreported if total proportions for gender or age did not equal between 98% and 102%

[b] Total includes all outbreaks where a single food commodity was identified