

Foodborne Disease Surveillance: Evaluation of a Consumer Driven Complaint System
and Development of Methods for Screening of Pathogens and Cluster Detection

A DISSERTATION
SUBMITTED TO THE FACULTY OF THE GRADUATE SCHOOL
OF THE UNIVERSITY OF MINNESOTA
BY

John Jiuhan Li

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

Craig Hedberg, PhD, Co-Advisor
Alan Lifson, MD, MPH, Co-Advisor

September 2010

© John Jiuhan Li 2010

Acknowledgements

Completing this dissertation would not have been possible without the guidance and support I received from numerous individuals. Firstly, I would like to thank my dissertation committee. I'm grateful to my adviser Craig Hedberg for listening to all my ideas, helping me focus my research, pointing me down paths for new discovery, and providing insightful feedback on my work. Thanks to my co-advisor, Alan Lifson, for bringing a different perspective and valuable advice. Thank you to Rich Maclehose for his great ideas, guidance through tricky statistical issues, and words of encouragement. To Jeff Bender for his thoughtful review and commentary. Thank you all for your guidance.

I also need to give great thanks to my friends. Thanks for lending an ear when needed and making this process a more enjoyable experience.

Lastly, I need to give much thanks to my wife, Ann. Without her support I would never have been able to complete this dissertation. She has stood by my side through this whole process giving me motivation when needed, a person to vent to when I was stressed, and taking my mind off of school when I needed a break.

Dedication

For my wife Ann, my daughter Charlotte, and my son Miles.

Abstract

Foodborne illnesses are common, an estimated 76 million cases occur in the United States annually, resulting in large annual costs from medical care, loss of productivity, and early mortality. Foodborne illnesses can result from food contaminated by various pathogens including bacteria, viruses, toxins, and parasites, making control and prevention a difficult task. The food supply is becoming more complex making preventing contaminated product from reaching the consumer more difficult. This makes early detection of outbreaks and removal of contaminated product essential to reducing the burden of illness from foodborne diseases.

The primary method for detection of foodborne disease is the use of public health surveillance. Surveillance for foodborne illnesses is typically carried out through pathogen specific surveillance or surveillance of consumer complaints of foodborne illness. In pathogen specific surveillance public health labs and physicians report cases of reportable illnesses to their local health department. National networks to for public health departments to share lab data and find potential cluster have been established. The drawback with this type of surveillance is that only a small percentage of those sick with a suspected foodborne disease will actually go see a physician. Another problem is that the lag time between onset of disease and reporting to a health department can be quite long, having to work its way through stool collection and testing prior to making it to the health department.

Complaint based surveillance based on the public directly reporting foodborne illnesses they suspect were caused by eating at a restaurant or a particular food product is another possible source of outbreak detection. However, complaint surveillance systems have not received the same level of attention as pathogen specific surveillance. Few studies have been carried out to examine the ability of complaint surveillance to detect outbreaks. Additionally, methods to efficiently use this data source have not been developed. This dissertation evaluated the use of complaint surveillance systems for outbreak detection and developed methods to better utilize complaints information.

In manuscript one, the complaint surveillance system in Minnesota was evaluated and characteristics of complaints that lead to detection of foodborne outbreaks were analyzed. It was found that complaint systems have the ability to detect outbreaks caused by a variety of pathogens. Case detection for foodborne disease surveillance in Minnesota happens through a multitude of mechanisms. The ability to integrate these mechanisms and carry out rapid investigations leads to improved outbreak detection. In manuscript two, a predictive model for *Salmonella* using complaint based surveillance data was developed. The predictive model was able to discriminate between *Salmonella* calls and other calls to the complaint system. A screening tool that can target specific pathogens for further investigation will help increase the utility of complaint data. In manuscript three, an algorithm was developed using cumulative sum methods to automatically flag weeks of unusually high call volume that could indicate clusters of illness. Two detection levels were used for our algorithm. The lower threshold found 6 out of 17

outbreaks in Hennepin County. The six outbreaks included 5 norovirus outbreaks and one large multi-state *Salmonella* outbreak. The algorithm found all outbreaks that were detected through multiple individual complaints. Sensitivity and specificity for the detection algorithm were 26% and 81% at the lower threshold when looking at all outbreaks. When focusing only on norovirus outbreaks the sensitivity and specificity were 63% and 84%. Manuscript four focused on the use of complaint base surveillance systems by local health departments in the United States. Eighty-one percent of health departments in the U.S. use some type of complaint based surveillance system. Complaint rates were found to be associated to outbreak rates with a pearson's correlation coefficient of 0.38 ($p=0.0004$). Those health departments with electronic databases appeared to have better ability to find outbreaks among complaint calls in their jurisdiction.

Complaint based surveillance systems have the ability to detect foodborne outbreaks and provide a necessary complement to pathogen specific surveillance. The use of both types of surveillance allows health departments to identify potential cases of illness and outbreaks from a variety of pathogens. Additionally, the majority of health departments have the infrastructure to use a consumer complaint system. However, most local health departments do not use these systems to their full potential to find outbreaks. This dissertation provides a framework for extending complaint surveillance use and application of more efficient ways to analyze incoming data to improve food safety in the United States.

Table of Contents

Acknowledgements	i
Abstract.....	iii
List of Tables	viii
List of Figures	x
Chapter 1. Introduction and Background	1
1.1. Overview	1
1.2. Foodborne Diseases	2
1.3. Epidemiology of Foodborne Illnesses.....	2
1.4. Norovirus.....	3
1.5. <i>Salmonella</i>	5
1.6. Preventing Foodborne Disease	5
1.7. Detection of Foodborne Diseases	6
1.8. Effectiveness of Surveillance	13
1.9. Data From Surveillance Systems.....	14
1.10. Current Methods in Outbreak Detection	15
1.11. Gaps and Limitations in Foodborne Disease Detection	17
1.12. Summary	19
Chapter 2. Data Collection and Study Methods.....	21
2.1. Data Source	21
2.2. Linking Complaints to Outbreaks.....	21
2.3. Variables	22
2.4. Bootstrap Internal Validation	23
2.5. Human Subjects and IRB Approval.....	25
Chapter 3. Manuscript 1	26
3.1. Introduction.....	28
3.2. Methods	30
3.3. Results.....	34
3.4 . Discussion	38
3.5. Tables.....	42
3.6. Figures	47
Chapter 4. Manuscript 2	48

4.1. Introduction	50
4.2. Methods	52
4.3. Results.....	55
4.4. Discussion	57
4.5. Tables.....	61
4.6. Figures	65
Chapter 5. Manuscript 3	67
5.1. Introduction.....	69
5.2. Methods	71
5.3. Results.....	75
5.4. Discussion	77
5.5. Tables.....	82
Chapter 6. Manuscript 4	86
6.1. Introduction.....	88
6.2. Methods	89
6.3. Results.....	91
6.4. Discussion	95
6.5. Tables.....	98
6.6. Figures	102
Chapter 7. Conclusion and Implication of Future Research.....	105
7.1. Key Findings	111
7.2. Strengths and Limitations	116
7.3. Implications for Future Research and Foodborne Surveillance Practices.....	119
Bibliography	123
Appendices.....	129

List of Tables

Table 1.1. Reporting steps in a surveillance system and cases reported to the next level ..9	9
Table 2.1: Variables in MDH surveillance database.....	22
 Manuscript 1	
Table 1: Source of case detection of confirmed foodborne outbreaks by etiology, Minnesota, 2000-2006.....	43
Table 2: Number of outbreaks in complaint database by the number of individual complaints associated with that outbreak in complaint system and median number ill per complaint	44
Table 3: Comparison of characteristics reported by the primary complaint caller by outbreak status.....	45
Table 4: Results of final logistic regression model examining variables associated with being an outbreak associated call.....	46
 Manuscript 2	
Table 1. Complaints from MDH complaint surveillance database from 2000-2008 with identified etiologies	61
Table 2. Univariate analysis of potential predictive variables to distinguish <i>Salmonella</i> calls from non- <i>Salmonella</i> calls	62
Table 3. Results of multivariate logistic regression to examine predictive variables a complaint being a <i>Salmonella</i> case	63
Table 4. Sensitivity, specificity, positive predictive value, and negative predictive value for 3 separate cutoffs for <i>Salmonella</i> , maximizing both specificity and sensitivity, maximizing specificity, and maximizing sensitivity	64
 Manuscript 3	
Table 1: Regression coefficients from autoregressive model of complaint data from Hennepin County, 2000-2006	82
Table 2: Outbreaks in Hennepin County in 2007 and flagged weeks from the cusum detection algorithm	83
Table 3: Sensitivity, specificity, positive predictive value, and negative predictive value for outbreak detection using the cusum algorithm at two detection levels	85
 Manuscript 4	
Table 1: Percentage of health departments with a consumer complaint surveillance system by population of health department jurisdiction	98
Table 2: Percentage of health departments that collect key information from complainants from those that said they had a complaint system	99
Table 3: Numbers of outbreaks, outbreaks detected by complaints, and rates of outbreaks and complaints in the past year by health agency by population category.....	100

Table 4: Summary of local health department use of consumer complaint surveillance systems 101

List of Figures

Figure 1	8
Manuscript 1	
Figure 1: Foodborne illness complaints by source of receipt and confirmed foodborne outbreaks by year, Minnesota, 2000-2006	47
Manuscript 2	
Figure 1: ROC curve of predictive model trained on original data	65
Figure 2: Count of complaint calls by <i>Salmonella</i> prediction by month	66
Survey Results	
Figure 1: Percentage of agencies that reported not having a foodborne outbreak in the past year by population size	102
Figure 2: Percentage of outbreaks found through complaint based surveillance systems with 95 percent confidence intervals	103
Figure 3: Comparison of median outbreak rates and outbreaks per complaint rates by population category and if the agency reported having an electronic database or not ...	104

Chapter 1. Introduction and Background

1.1. Overview

Foodborne disease continues to be a major problem in the United States and around the world [1]. An estimated 76 million cases of foodborne disease occur annually in the United States at a cost of up to 30 billion dollars [2-4]. In industrialized nations up to 30% of the population has been reported to suffer from foodborne disease annually [5]. The World Health Organization estimates that globally in 2005 there were 1.8 million deaths from diarrheal disease. The majority of these were due to contaminated food or water. Foodborne diseases are caused by a multitude of pathogens making detection and prevention difficult. Our food system is growing ever more complex. As our marketplaces become more global and food distribution networks become wider the challenge of controlling foodborne diseases will also increase [6]. The average grocery store carries over 30,000 food items, many of which are ready to eat foods made of a number of ingredients making identification of contaminated items difficult [7]. In this changing food environment new approaches to foodborne illness detection need to be explored. Currently, detection of foodborne illness is largely based on pathogen specific surveillance. This depends on ill individuals going to seek medical attention and having enough cases that common exposures can be linked. Due to the complexity of the foods we eat this can be problematic. New methods using surveillance that allows direct reporting of illness, such as a consumer complaint driven surveillance system, could potentially aid in foodborne disease investigations.

1.2. Foodborne Diseases

Foodborne diseases are commonly defined as diseases that are caused by agents or pathogens that enter the body through ingestion of food. Foodborne diseases are caused by a variety of agents, including bacteria, viruses, parasites, and toxins. During 1998-2002, a total of 6,647 outbreaks were reported to the Centers for Disease Control and Prevention in the United States. The majority of these outbreaks were of unknown etiology, only 33% had confirmed etiologies [8]. Norovirus and *Salmonella* were the leading pathogens in outbreaks with known etiologies causing 10% and 9% of illnesses respectively. Many of the unknown cases can also be attributed to Norovirus. These numbers are likely to be lower than the true number of outbreaks since they depend on reporting of states and local agencies to the CDC. Food can be contaminated through various mechanisms including, handling of food by an ill food worker, contamination of product on the farm, or contamination sometime during the production process [9].

1.3. Epidemiology of Foodborne Illnesses

Emergence of foodborne illnesses can be affected by many things including changes in human behavior, food technology, economic development, commerce, and microbial adaptation [10]. Increased international travel and evolving eating behaviors have changed the patterns of disease. Change at any level of the food chain that can affect rates of foodborne disease.

Characteristics of foodborne illnesses vary from pathogen to pathogen. Most show some type of seasonality, however, the seasonality of pathogens may be different. The common factor in all disease is that they are caused by ingestion of a contaminated food product or close contact with a sick individual. Everyone is at risk for foodborne diseases. However, complications due to illness are most often found in the elderly, young children, and immunosuppressed [11, 12]. Other risk factors reported are international travel, eating outside of the home, and poor food handling practices, and close contact with an ill individual [13-15]. Norovirus is the most common cause of acute gastrointestinal illness. *Salmonella* is the most common bacterial cause of foodborne outbreaks. Specifics on these two pathogens will be presented since they represent two very different organisms that any foodborne disease surveillance system will need to be able to detect.

1.4. Norovirus

Norovirus causes outbreaks in many settings including restaurants, schools, hospitals, cruise ships, and nursing homes [16, 17]. Norovirus induced illness usually includes both vomiting and diarrhea. Symptoms can also include nausea, abdominal cramps, and systemic symptoms such as malaise, myalgia, chills, and headaches [18]. Low grade fever is also present in half the cases. Incubation time is usually 24 to 51 hours. Illness normally resolves in 24-72 hours, but severe disease can be experienced by young children, the elderly, and immunosuppressed populations. Norovirus was long thought to have a strong seasonal component with a rise in cases in the fall and winter [19, 20].

However, more recent studies on norovirus outbreaks in the United States have found a rising number of outbreaks past the peak season [16].

Norovirus causes a large amount of disease for several reasons. One is that the infectious dose appears to be quite low, as few as 10 virus particles [21]. Secondly, prevention of spread of disease is incredibly difficult. Norovirus is extremely hardy, being resistant to freezing, heating, exposure to low levels of chlorine, and detergent based cleaners [22]. The virus can shed for long periods after symptoms pass [23]. Lastly, there is no treatment for norovirus. Prevention is the only tool for limiting disease. Early detection of outbreaks could help limit the spread of illness by implementing enhanced sanitation procedures and reducing contact with ill individuals.

1.4.a. Risk Factors for Norovirus

The most important risk factor reported for norovirus is contact with an ill person [24]. This can also include preparation of food by a sick food handler. Norovirus is an important disease in nosocomial infections [25]. Some study has been done looking at institutional risk factors for norovirus. One study looked at attributes of hospitals that may make them particularly vulnerable to norovirus outbreaks; geriatric wards were much more likely to experience norovirus outbreaks, with a hazard ratio of 2.6 when compared to general medical care units [26, 27]. Also the size of the institution affected

the rate of norovirus outbreaks, with a hazard ratio of 1.22 for every 10 additional beds [26].

1.5. *Salmonella*

Salmonella is the most common bacterial cause of foodborne illness outbreaks in the United States [8]. It has over 2500 serotypes that often have different epidemiology and reservoirs. *Salmonella* often presents as acute gastroenteritis which resolves on its own, however serious outcomes can be experienced in immunosuppressed individuals, elderly, and the young [28]. *Salmonella* also shows some seasonality, with peaks in cases in summer months [29]. Surveillance for *Salmonella* largely depends on passive laboratory reporting [29].

1.5.a. Risk Factors for *Salmonella*

Salmonella has numerous serotypes that may have differing risk factors. For the most common serotype, *Salmonella enteritidis*, risk factors have been found to be international travel, consumption of chicken outside the home, undercooked eggs, and contact with birds and reptiles [30]. Egg consumption has also been shown to be a risk factor for other *Salmonella* serotypes [31]

1.6. Preventing Foodborne Disease

Prevention of foodborne disease is a difficult task, which requires work at all levels of the food supply. Improved food safety practices such as hazard analysis and critical control

point (HACCP) have been widely implemented in the food industry [32]. However there has not been a corresponding drop in foodborne disease incidence [33]. The food chain is complicated and pathogens can be introduced into food through numerous sources and pathways. Other avenues for preventing disease are also difficult. Treatments are not available for many infections and educating the consumer is not always sufficient. Education is important but many illness events are out of control of the consumer. In addition the variety of risk factors and differences between pathogens make education difficult [34]. The other main tool to preventing disease is early detection and abatement of sources of contamination [35]. Detecting disease also allows for the detailed study of disease process to aid in making the food supply safer and preventing future cases of illness.

1.7. Detection of Foodborne Diseases

Surveillance systems are the cornerstone of outbreak detection by public health agencies. Public health surveillance has been defined as the ongoing systematic collection, analysis, interpretation, and dissemination of health data for the purpose of preventing and controlling disease, injury, and other health problems [36]. In the context of foodborne diseases, the goal of surveillance is the detection of outbreaks for two purposes; to prevent further spread of disease, and to gather information on possible contributing factors to outbreaks for further study. Most surveillance is passive in nature, meaning that it waits for cases or illness to be reported to the system. There are two

primary methods of surveillance for foodborne diseases, pathogen specific surveillance and surveillance based on consumer complaints.

1.7.a. Pathogen Specific Surveillance

Efforts to standardize pathogen specific surveillance across the United States have been carried out. Standard methods for identifying and subtyping pathogens are regularly disseminated. Public health labs also are able to share results. Pulesenet, a platform for sharing results of lab testing nationwide and internationally has been established, with participation from all 50 states and several foreign countries [37]. Pulesenet was created in 1996 as a molecular surveillance network for foodborne infections [38]. Results from pathogen subtyping are shared through Pulesenet to allow detection of possible clusters of illness. This type of data sharing has led to the detection of numerous outbreaks [39].

Pathogen specific surveillance is dependent on the reporting of notifiable diseases by laboratories and clinicians to local health departments. The CDC in collaboration with public health officials at various levels of government determines which diseases should be nationally notifiable. Each jurisdiction can also add to this list if they deem it necessary. Any instance of a laboratory or clinician finding these pathogens is reported to the local health agency. As of 2006, only five diseases commonly associated with food were nationally notifiable; Botulism, *Cryptosporidiosis*, *Giardiasis*, *E. coli* O157:H7, and *Salmonellosis* [40]. In the state of Minnesota, 8 foodborne diseases are

reportable, the 5 mentioned previously and; *Campylobacteriosis*, *Cholera*, and *Listeriosis*. Physicians also may report clusters of illness that are not reportable but unusual.

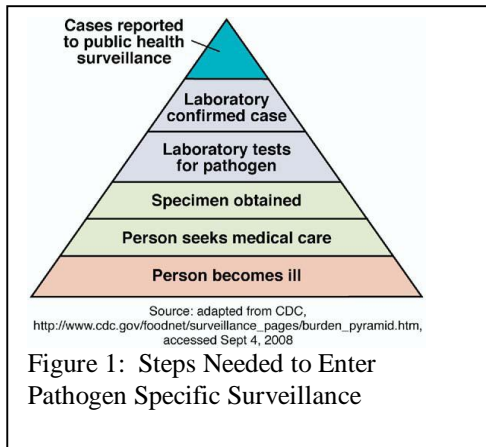


Figure 1: Steps Needed to Enter Pathogen Specific Surveillance

Pathogen specific surveillance has its drawbacks in the limited number of diseases that are reportable and issues with underreporting. Figure 1 illustrates how only a small fraction of people who become ill with a foodborne disease are finally picked up in pathogen specific surveillance. Population based studies have

also shown that those ill with foodborne disease are not likely to seek medical attention [41]. Various estimates have been put forward of the underreporting of foodborne disease [4]. Underreporting often differs by pathogen. Those that cause milder symptoms are more likely to be underreported.

1.7.b. *Salmonella* Specific Surveillance

One of the most developed pathogen specific surveillance system is for *Salmonella*. In 1962, a surveillance system specifically for *Salmonella* was started. It has since grown to encompass several different surveillance mechanisms. The national system of *Salmonella* surveillance includes laboratory serotyping data, reports of foodborne outbreaks, and lab monitoring of anti-microbial resistance [42]. This system is very

robust, but it has several drawbacks. The data is not linked well, meaning it is difficult to connect cases in the disparate systems. Additionally, it suffers from large amount of underreporting.

Table 1.1: Reporting Steps in a Surveillance System and Cases Reported to the Next Level

Surveillance Step	Percentage Reported
1. Patient infected	-
2. Patient ill	45%
3. Patient sees doctor	45%
4. Doctor obtains culture	45%
5. Laboratory confirms organism	71%
6. Laboratory reports to the health department	50%
7. Health department reports to the CDC	83%

Adapted from Table 4 in article by Chalker et al. [43]

Table 1 illustrates the amount of underreporting at each level of a surveillance system for *Salmonella* infection. The percentage reported is an estimate of the number of cases that are reported to the next level of surveillance. Most cases are lost in the early stages of a surveillance system where patients decide to seek healthcare, and a physician obtains a stool culture. Surveillance that allows the public to direct report to health departments may be able to find more cases of illness.

1.7.c. Consumer Complaint Systems

Consumer complaint surveillance systems provide another surveillance mechanism for foodborne illnesses. These types of systems allow the general public to directly contact public health agencies to report eating establishments or products that they suspect caused their disease. The determination of the cause of disease and if it is food related is completely up to the individual complainant. Very little study has been carried out on these types of surveillance systems. Nothing is currently known on the usage of this type of surveillance in the United States. There are no uniform standards in type or method of data collection. It is also unknown the extent of use of such systems in the United States. Complaints systems can range from the simple anonymous reporting of a suspect eatery, to detailed collection of disease information, suspected eatery, and food history. Additionally, how the data is used varies. Some jurisdictions keep paper copies of reports for possible follow up by sanitarians, to computerized databases that can be searched for common exposures.

Complaint system surveillance has been cited as one of the best methods for detecting non-reportable pathogens and new and emerging agents [44]. Only one study has been carried out focusing on complaints system. They found that consumer complaints data is likely contains information about many foodborne outbreaks [45]. The uncertainty of the effectiveness of complaints systems stems in large part on the lack of study in this area and the varied state of complaints surveillance. Jurisdictions that collect little information will have a much harder time making linkages that lead to discovery of

potential outbreaks. Additionally, a major drawback of a complaints system is the volume of calls received and difficulty in identifying calls to investigate further.

1.7.d. Complaints Reporting Behavior

No study has been carried out specifically investigating callers to complaints systems. However, one study examined public beliefs about foodborne illnesses [41]. In the study, the authors explored reporting behaviors of individuals that suspected they had a foodborne illness. Of those who were ill only 8% reported their illness. Those that suffered from vomiting or missed work were more likely to report their illness. When individuals were asked why they suspected a food establishment caused their illness most cited timing of illness and others sick in their party as reasons they suspected the restaurant. However, when asked about the timing of illness most reported that they had eaten within 5 hours of getting ill. This is shorter than most incubation times for foodborne illness, suggesting that individuals have difficulty in determining the cause of their illness.

1.7.e. Minnesota Department of Health Foodborne Disease Investigation

In Minnesota, the model developed for foodborne illness outbreak detection and investigation involves centralized interviewing of ill individuals at MDH. Outbreak investigations are done primarily by MDH with support from local health departments. A few larger local health departments in Minnesota conduct their own investigations, but

always with input and collaboration from MDH. Data on foodborne illness cases are stored in a central database at the state level. The database indicates if information on the case was received through pathogen-specific surveillance or not. Cases not received through pathogen-specific surveillance are received through provider reports of illness clusters, reports from institutions, reports of complaints from restaurants, or, in the majority of cases, direct consumer complaints to the MDH foodborne illness hotline.

The Minnesota Department of Health (MDH) has a very robust complaint system. The public can report illness to this system through a toll free hotline number, 1-877-Food-ill. The hotline was started in 1998 and receives calls from residents throughout the state. Approximately, 800 calls are received by the annually. Callers are instructed to call in if they believe a specific food product or establishment caused their illness. The instructions on calling can be found online and in brochures. Some calls are also received by local health departments. Local health departments are asked to either forward complaint callers to MDH or fax a completed case report form. Complaints are also occasionally forwarded to MDH from food establishments. A standard intake form is used to record complainant information (Appendix A). Questions cover demographics of the caller, illness information, suspected food product or establishment, and names and contact information for other members of the dining party (if applicable). If illness is limited to members of a single household, a 4-day food history is obtained, with focus on meals eaten outside of the home. If illness is reported among members of multiple households, information is taken only on meals common to members of the different

households. All information collected is entered into the MDH complaint database.

Complaints involving multiple households, instances of multiple independent complaints about the same restaurant, and reports of clusters of illness are evaluated by the MDH foodborne illness supervisor, and outbreak investigations initiated with the appropriate state or local health agencies.

1.8. Effectiveness of Surveillance

There are several key criteria that are used in determining the effectiveness of a complaints system. The criteria all are related to the main goal of a complaints system which is detecting illness. Specific criteria are; is the system acceptable to health agencies and users, it is flexible enough to adapt to changing needs, does it find disease, is it timely, and is the cost reasonable [46]. Every system has its strengths and weaknesses and by pairing up different sources of data such as lab and complaints data the overall surveillance system could be enhanced. No one system is enough to address the complex issues in foodborne illness detection.

As mentioned earlier, risk factors for disease and patterns of illness may differ from pathogen to pathogen. This is the most difficult aspect in using surveillance data. Many data sources for surveillance systems focus entirely on a specific pathogen. An ideal foodborne illness surveillance system would have to be able to account for many pathogens and trigger investigation prior to confirmation of etiology. *Salmonella* surveillance systems have been in existence for decades. However, they do not allow the

ability for detection of other types of pathogens. They also are heavily dependent on cases seeking medical attention.

1.9. Data From Surveillance Systems

The type and amount of data available differs with different surveillance mechanisms. Pathogen specific surveillance is able to provide confirmed etiologies and usually demographic information on the case. Other systems often cannot provide pathogen information, but may be able to provide information on common exposures, illness information, and data on other ill individuals. Follow-up on each case depends on the level of urgency each health department treats foodborne illnesses and available resources.

Ideally, data from surveillance systems is reported and shared among neighboring health jurisdiction and higher levels of health jurisdictions. The CDC asks that health departments report foodborne outbreaks to them through their EFORS electronic reporting system. However, in practice this doesn't always happen. The estimated underreporting presented in Table 1 were for *Salmonella* only, however, underreporting is likely for all pathogens [4, 47]. It is probably even worse with nonreportable illnesses and those illnesses that cause milder symptoms such as Norovirus [48]. Surveillance reporting is also complicated by home state laws. Each state and local jurisdiction has their own methods for investigation and reporting of foodborne disease. In some States, investigations and data are centralized at the State level. In other States local health

departments are in charge of their own investigations and data and resources are decentralized. There are also large time lags as information is passed up the chain of reporting. The greater the time from the time of illness onset, the less useful the data becomes in mitigating ongoing outbreaks. By the time information is collected at the CDC, any studies are generally retrospective in nature and unlikely to detect ongoing outbreaks.

1.10. Current Methods in Outbreak Detection

Detection of outbreaks among cases of foodborne illness is dependent on being able to find a common cause of illness among individuals. This always involved follow up with reported cases. Health department personnel can follow up with reported cases to find out more about their illness. Most health departments do not have the resources to investigate every case so through their own procedures they decide which ones should be investigated. This procedure is often a judgment by an experienced public health practitioner. Due to the subjective nature of decision making, the ability to detect outbreaks is highly dependent on the culture of each health agency. Those that treat cases with more urgency will undoubtedly find more illness. Using a person's judgment will also become more difficult if call volume is large, or if the United States moves to a more regionalized foodborne disease surveillance system. On a higher level CDC has implemented automated systems to help detect clusters of disease in regards to *Salmonella* surveillance. These systems take data from laboratory surveillance and outbreak reporting to try to find clusters of illness.

Detection of outbreaks through linking numerous cases works well with our traditional idea of an outbreak. Traditional outbreaks are those that follow the common perception of an outbreak, meaning large number of people ill in a small area who perhaps all attended the same event. However a new scenario of outbreaks is becoming much more common. In this case illness is low level and widespread [35]. They may only affect a small proportion of those exposed making detection very difficult.

Symptoms of foodborne illnesses are often non-specific on the individual level but on the aggregate level they have been used to examine etiology of outbreaks [49]. Use of clinical symptoms has been suggested as a tool in investigation of foodborne outbreaks. One study examined the ability of clinical profiles to correctly classify outbreaks of known etiology. Sensitivities for clinical profiles in the study were found to range from 62% to 78% [50]. Many outbreaks are of unknown etiology due to the lack of testing of individuals. Tools to determine etiology are useful in studying past outbreaks. These profiles in the context of complaints could potentially be used prospectively to help screen for pathogens.

Syndromic surveillance also uses symptoms of cases but in a slightly different way than profiling of outbreaks. Syndromic surveillance is the use of symptom information to find clusters of disease. Syndromic surveillance has also been suggested for use in outbreak detection [51]. In the area of bioterrorism use of symptoms in the form of syndromic

surveillance have been suggested as a tool for detecting possible attacks [52]. Syndromic surveillance has also been used to look at foodborne illness [53]. In most studies syndromic surveillance was based on hospital discharge records. This provides a large number of people to study, but it also has a major limitation in that a case needs seek healthcare. However its main drawback is that symptoms for foodborne illness are rather non-specific. It would be hard to distinguish cases in a large pool of people with other illnesses. However, using symptom information for people that have self identified themselves with a foodborne illness may be more promising.

CDC uses temporal correlation to help find clusters of *Salmonella* in laboratory reported subtypes. The system is called SODA and uses data from Pulsenet which links laboratories across the United States [42]. This system uses cumulative sum methods to detect unusually high numbers of reported *Salmonella* sub types reported from laboratories. Cumulative sum methods look for cases accruing over time and triggers alerts for unusually high numbers of cases. SODA has been effective in detecting a nationwide *Salmonella* outbreak caused by mangoes [54].

1.11. Gaps and Limitations in Foodborne Disease Detection

Pathogen specific surveillance is limited by severe underreporting and small number disease that are reportable. Pathogen specific surveillance also depends on physicians ordering stool tests and reporting of notifiable illnesses. In a recent study of emergency room physicians only 38% ordered stool cultures for patients suffering from acute

diarrheal illness [55], illustrating the importance of exploring other surveillance systems that can be used in conjunction with laboratory surveillance to detect more cases of foodborne illnesses. One previous study has been carried out looking at complaints systems for foodborne illness outbreak detection [45]. The authors found that the complaint system likely contained many calls linked to possible outbreaks. However, their study was limited in that they could only analyze one year of data in the city of San Francisco, and had limited number of investigated outbreaks. They also could not compare data from complaint surveillance to laboratory surveillance.

Average lag times between illness onset and reporting to the health department have been found to range from 2 to 3 weeks days [56, 57]. This lengthy lag time makes stopping outbreaks as they are occurring extremely difficult. One survey of health departments found that delayed notification of an outbreak was a major reason that they were not investigated [58]. Reducing lag times may encourage more investigations and increase the urgency of investigation. It will also allow for the prevention of more cases of illness.

Statistical methods to detect clusters of illness have been developed for laboratory based surveillance for foodborne illness and surveillance of other illnesses, but no work has been done with consumer complaints data. The difficulty with using consumer complaint detecting the proper complaints to investigated among the often high volume of complaints. Application of statistical methods to help identify calls is needed to increase the utility of complaint surveillance. Combining the multiple streams of data, such as

complaint data and laboratory surveillance, can give a more complete picture of foodborne disease patterns.

1.12. Summary

The future control of foodborne diseases will depend on early detection of disease and a cohesive investigation system [46]. Timeliness of outbreak detection will generally be improved by improving detection of individual cases [59]. It is critical to find surveillance systems that can be used in conjunction with laboratory surveillance to detect cases. Complaint systems are perfectly positioned to fill this role. Current laboratory surveillance methods have their drawbacks that complaint systems may be able to fill. Using data from both types of surveillance could greatly aid in early detection of foodborne illness. Evidence of the utility of a complaint system could greatly help in increasing its utilization across the country.

Laboratory surveillance also does not do well in detection of some pathogens, in particular those that are not routinely tested. One of those pathogens is norovirus, the leading cause of acute gastrointestinal illness. Those suffering from norovirus often don't seek medical attention meaning no stool samples are ever tested. Limiting outbreaks of norovirus depends on other means of finding cases.

Complaint based surveillance is used in many jurisdictions in one form or another.

However, their most effective use is not well established. By examining factors that help

detect outbreaks model systems can be developed and put into use. This would not be overly burdensome since health agencies run surveillance programs as part of usual public health practice. Early detection of more cases of foodborne illness would lead to earlier detection of outbreaks and prevention of further illness. Finding more cases would also give the health departments the ability to detect smaller outbreaks than may now go undetected.

Developing better methods of screening incoming data will also assist in early detection of outbreaks. Being able to screen complaints data for likely cause of illness would be an important tool to better use incoming data. Not all jurisdictions investigate norovirus, in this case complaints surveillance would be vastly improved if there was a way to reduce calls that are suspected norovirus. Additionally developing automated flagging systems to identify clusters of illness are necessary to further the use of complaint surveillance data.

The goal of this dissertation was to improve food safety in the United States by development of improved methods for foodborne illness surveillance. This was done through four manuscripts that studied various aspects of foodborne disease surveillance, by examining a possibly underused source of data and development of new methods for detecting clusters of illness. The four manuscripts together fill the current gap in knowledge in the effectiveness of using complaint based surveillance for foodborne illnesses and methods to best utilize data from these surveillance systems.

Chapter 2. Data Collection and Study Methods

2.1. Data Source

The primary source of data for this dissertation was the Minnesota Department of Health's (MDH) foodborne illness surveillance database. Data from 2000-2006 was used in manuscript one to evaluate the effectiveness of MDH's complaint based surveillance system. Data from 2000-2007 was used in manuscripts two and three. As discussed previously the MDH complaints line take complaint calls from the general public in the State of Minnesota. Callers report a food product or establishment that they feel caused their foodborne illness. A standardized set of questions are asked and entered into a database at the state. MDH stores each year's database separately. Deidentified complaints databases for the span of the study were obtained from MDH. Data on confirmed foodborne outbreaks in the form of annual gastroenteritis summaries published MDH were also obtained. These summaries were used to determine the method of detection of each outbreak during the study period.

2.2. Linking Complaints to Outbreaks

Only a small number of complaints are investigated and found to be part of a confirmed foodborne outbreak. Complaints involving multiple households, instances of multiple independent complaints about the same restaurant, and reports of clusters of illness are evaluated by the MDH foodborne illness supervisor, and outbreak investigations initiated with the appropriate state or local health agencies. Further investigation of these illnesses

could lead to identification of a foodborne outbreak through confirmation of a pathogen through laboratory testing or confirmation of a common exposure. Standard practice up until 2007 was to mark the first complaint as an outbreak if multiple complaints were involved. To link calls to outbreaks the yearly MDH gastroenteritis outbreak summaries were used in conjunction with outbreak databases. The outbreak summaries are published for each year and contain a short description of every outbreak, often including dates of complaint calls. This information was used to match date of complaints and location of outbreaks to determine other outbreak associated calls.

2.3. Variables

The foodborne illness database contains all the information collected from callers as well as results from any lab tests that may have been done. This research did not use all fields of collected information. The variables in the following table were used in this analysis

Table 2.1

Variable	Description
Daily	Source of Complaint, direct from consumer or laboratory surveillance
Form Date	Date complaint was taken
Onset Date	Onset date of illness
Suspect Date	Date and time of the suspected meal that caused illness
Symptoms	5 symptoms, diarrhea, vomiting, cramps, fever, bloody stool, coded as a yes no,
Outbreak	If complaints (or one in a group) was associated to an outbreak
County	County of residence of caller
Establishment 1-4	Establishments that caller had eaten in the 4 days previous to illness
Suspected Establishment	Establishment that the caller suspect made them sick

Number Ill	Number of people ill in caller's party
Age	Age of caller
Gender	Gender of caller
Physician Visit	If caller went to see a physician due to illness
Organism	Etiology of a caller's illness if a stool sample was tested

Some variables will be used to generate new ones for analysis. Lag time between onset date and form date will also be created. Also, suspected incubation times of a caller's illness were calculated using the reported onset of illness and suspected meal time. This is likely not the true incubation period of the illness since it is not know if the suspected meal time is the correct cause of illness.

2.4. Bootstrap Internal Validation

Bootstrap methods were used in manuscript two to internally validate the predictive model that was created. More details of the method used are provided in this section. The bootstrap is a data-based simulation method used for statistical inference based on resampling with replacement from the original data [60]. In the early 1980s, Efron et. al. led the development of bootstrap methods with several papers discussing the wide applicability of use of the method in determining standard errors, confidence intervals, model selection, regression analysis, and cross validation [61-65].

Bootstrap methods have also been used to internally validate prediction models. Prediction models often suffer from overfitting. Performance measures of these models such as the area under curve (AUC) used in manuscript two are often optimistic. The

reason being is that these measures are computed through self-prediction, in other words they use the same data to create the model and obtain predictions. In this case it is recommended to take measures against overfitting, often called validation. There are various validation methods including split sample validation, which as it is called splits the data in to two datasets to compare results. However, this does not work well with smaller sample sizes. A few sample reuse methods such as cross validation, and bootstrap validation do not require the splitting of samples. Bootstrap validation was chosen for this dissertation because it has been found to provide the most stable estimates of optimism correction out of several different interval validation methods [66].

Bootstrap validation requires creating a certain number of bootstrap samples and running the prediction model on each one. The purpose of this is to correct the overly optimistic AUC. In the analysis presented in manuscript two, 1000 samples were created by randomly sampling from the original data with replacement. The predictive model is fit on each of these training samples finding the AUC for each one. The parameter estimates from each trained model is then also fit on the original data finding the resulting AUC. There are several ways that the bootstrap estimates can be used to correct the AUC. One method is to take the difference in the average difference in the AUCs and subtracting this from the AUC calculated from the original model. An improvement over this is the .632 method.

In this study the .632 method was used to convert these two AUCs into an adjusted AUC. This .632 bootstrap estimator was proposed by Efron as a better method to determine an unbiased error rate [60, 66]. The .632 method multiplies the mean AUC from the 1000 bootstrap samples by .368 and adds it to .632 times the mean AUC of the trained model applied to the original data. The factor of .632 is used because it has been shown that bootstrap samples used in calculating the error are on average $1/0.632$ further away from the average [60]. This method corrects for this in calculating an adjusted error providing a better estimate than simply adjusting by the average difference. The adjusted AUC was reported and compared to the AUC calculated from the original data.

2.5. Human Subjects and IRB Approval

Complaint Surveillance data used in this study comes from existing databases at the Minnesota Department of Health. Data was de-identified prior to analysis. Original data did not go through IRB since it was collected by routine public health surveillance. Due to the low risk in the study and the fact that no personal identifiers will be used, this secondary data analysis has been deemed exempt from IRB review by the University of Minnesota Institutional Review Board (Study Number 0802E27412). The University of Minnesota IRB was also consulted regarding the survey of local health departments. It was decided that the survey did not meet the regulatory definition of research with human subjects since individuals were asked about the policies of their health department and not about themselves.

Chapter 3. Manuscript 1

Title: Evaluation of a Statewide Foodborne Illness Complaint Surveillance System,
Minnesota, 2000-2006

Foodborne outbreaks are detected by recognition of similar illnesses among persons with a common exposure, or by identification of clusters through pathogen-specific surveillance. Pulsenet has created a national framework for pathogen-specific surveillance, but there has been no comparable effort to improve surveillance of consumer complaints of suspected foodborne illness. The purpose of this study was to characterize the complaint surveillance system in Minnesota and evaluate its ability to detect outbreaks.

Minnesota Department of Health (MDH) foodborne illness surveillance data from 2000-2006 were analyzed for this study. During this period consumer complaint surveillance led to detection of 79% of confirmed foodborne outbreaks. The majority of norovirus outbreaks were detected through complaints. Complaint surveillance also directly led or contributed to detection of 25% of salmonellosis outbreaks. 81% of complainants did not seek medical attention. The number ill in a complainant's party was significantly associated with a complaint ultimately resulting in identification of a foodborne outbreak. This was related to a complainant's ability to self identify a common exposure, it also was likely related to the process by which MDH chooses complaints to investigate. There was a significant difference in incubation periods of complaints that were outbreak-associated (median, 27 hours) vs. those that were not outbreak-associated (6 hours, $p < 0.001$). Complaint systems have the ability to detect outbreaks caused by a variety of pathogens. Case detection for foodborne disease surveillance in Minnesota happens through a multitude of mechanisms. The ability to integrate these mechanisms and carry out rapid investigations leads to improved outbreak detection.

3.1. Introduction

Foodborne illnesses affect an estimated 76 million people and cost up to 30 billion dollars a year in the United States [2, 4]. Many illnesses stem from food eaten outside of the home in outbreak and non-outbreak situations [67-70]. The complexity of the food supply increases the number of control points needed to prevent foodborne illness, increasing the possibility of contaminated product reaching the consumer. Thus, early detection of outbreaks is critical to reducing the burden of foodborne illnesses by abating the source of contamination and identifying common contributing factors that could be controlled to prevent future outbreaks [35].

A foodborne outbreak is defined as the occurrence of two or more cases of a similar illness resulting from the ingestion of a common food. However, there are no standard methods for detecting and reporting these events [8]. In the United States, state and local health departments are responsible for investigating foodborne illnesses. Detection of foodborne outbreaks typically occurs through one of several ways: identification of a cluster of cases through pathogen-specific surveillance, investigation of consumer complaints of suspected foodborne illness, or report of clusters of illness by health care providers or institutions.

Pathogen-specific surveillance is based on reports of diagnostic laboratory tests or individual case reports from doctors' offices to public health agencies. The use of

molecular subtyping by pulsed-field gel electrophoresis and the sharing of subtype patterns on a national basis through PulseNet has greatly increased the sensitivity of outbreak detection for *Salmonella* and *E. coli* O157:H7 [39, 71, 72]. A drawback of pathogen-specific surveillance is that it is dependent on an ill person seeking health care. However, individuals suffering from enteric illnesses often do not seek healthcare and many of those who do are never tested. One estimate suggests that for every *Salmonella* case reported to the health department through lab surveillance 38 are unreported [4, 43]. Pathogen-specific surveillance also suffers from long lag times. The interval from illness onset and report to a health department has been estimated to be anywhere from 2 to 3 weeks [56, 57]. Lengthy lag times are problematic since a survey of local and state epidemiologists showed delayed notification as a major barrier to investigating foodborne disease cases [58].

Investigation of consumer complaints of suspected foodborne illness also provides information for public health surveillance systems. Complaint systems are the only way to detect outbreaks caused by non-reportable pathogens and emerging pathogens [44]. Non-reportable illnesses include many of the common causes of foodborne illness, such as norovirus, *Clostridium perfringens*, and less frequently recognized but potentially emerging foodborne pathogens such as enterotoxigenic *E. coli*. Direct reporting of exposures at the time of illness may also allow complaint systems to detect outbreaks faster than pathogen-specific surveillance.

Little work has been done to systematically investigate how effective complaint systems are for detecting foodborne illness outbreaks and the type of information that should be collected. In this study we analyzed the Minnesota Department of Health (MDH) foodborne illness complaint system. The objectives of the study were: to quantify sources of case detection for confirmed foodborne outbreaks, compare characteristics of outbreak and non-outbreak-associated complaints, and to evaluate what information regarding the complaint should be collected to allow detection of an outbreak.

3.2. Methods

In Minnesota, the model developed for foodborne illness outbreak detection and investigation involves centralized interviewing of ill individuals at MDH. Outbreak investigations are done primarily by MDH with support from local health departments. A few larger local health departments in Minnesota conduct their own investigations, but always with input and collaboration from MDH. Data on foodborne illness cases are stored in a central database at the state level. The database indicates if information on the case was received through pathogen-specific surveillance or not. Cases not received through pathogen-specific surveillance are received through provider reports of illness clusters, reports from institutions, reports of complaints from restaurants, or, in the majority of cases, direct consumer complaints to the MDH foodborne illness hotline.

MDH has run a statewide foodborne illness complaint hotline since 1998. Residents of Minnesota are encouraged to call a toll-free number if they suspect they have a food-related illness. Individuals also can e-mail a complaint to MDH via the MDH web-site. Local health departments are asked to either forward complaint callers to MDH or fax a completed case report form. Complaints are also occasionally forwarded to MDH from food establishments. A standard intake form is used to record complainant information. Questions cover demographics of the caller, illness information, suspected food product or establishment, and names and contact information for other members of the dining party (if applicable). If illness is limited to members of a single household, a 4-day food history is obtained, with focus on meals eaten outside of the home. If illness is reported among members of multiple households, information is taken only on meals common to members of the different households. All information collected is entered into the MDH complaint database. Complaints involving multiple households, instances of multiple independent complaints about the same restaurant, and reports of clusters of illness are evaluated by the MDH foodborne illness supervisor, and outbreak investigations initiated with the appropriate state or local health agencies.

MDH staff conducts standard interviews on all cases detected through pathogen-specific surveillance with a variety of reportable pathogens, including *Salmonella*, Shiga toxin-producing *E. coli*, *Shigella*, *Campylobacter*, *Listeria monocytogenes*, *Vibrio* spp., *Yersinia* spp., *Cryptosporidium*, and *Cyclospora*. All restaurants that cases reported eating at within the 7 days prior to illness onset are also entered into the MDH complaint database.

Data from the MDH complaint database and summaries of confirmed gastroenteritis outbreaks from 2000-2006 were used for this study. A confirmed foodborne outbreak was defined as an incident in which two or more persons experienced a similar illness after ingestion of a common food or meal, and epidemiologic evaluation implicated the meal or food as the source of illness. Outbreaks were characterized by etiology and method of outbreak identification. Etiology was primarily determined through laboratory confirmation of the pathogen through the outbreak investigation. In some instances laboratory confirmation was not possible. When this occurred, the etiology was epidemiologically classified as either a diarrhea toxin or vomiting toxin when characteristics of an outbreak met the criteria given by Hall et al. [49]. The etiology was classified as suspected norovirus when characteristics of the outbreak met the criteria laid out by Kaplan [73] and Hedberg et al. [50]. Outbreaks were classified as being detected through consumer complaint, report of a cluster of illnesses from a healthcare provider or institution, or pathogen-specific surveillance. Consumer complaints were further broken down into complaints about an event (e.g., wedding reception, party, conference, catered workplace gathering, etc.) or a food establishment. Complaints about a food establishment were categorized based on whether the complaint was initially made directly to a health department or to the establishment. Pathogen-specific surveillance was broken down further into strictly laboratory-based surveillance (i.e., receipt of isolates by the MDH Public Health Laboratory, with subsequent confirmation of species,

serotyping, and molecular subtyping), and healthcare provider report of a reportable illness before the isolate has received and typed by MDH.

Complaints that were associated with confirmed foodborne outbreaks that were detected solely by a consumer complaint or in conjunction with a consumer complaint were identified, and classified as outbreak-associated. The number of independent complaint calls per outbreak was tabulated. Symptoms and healthcare-seeking behavior of callers were also examined. The timeliness of complaint-based surveillance was compared to that of laboratory-based surveillance. Timeliness was defined as the number of days between illness onset and report to the public health department [46]. Median times were compared using a Wilcoxon-Mann Whitney test. Nine observations were omitted from analysis due to having lag times surpassing 100 days. SAS version 9.2 was used for all analyses.

Univariate analyses were carried out to examine which characteristics of a complaint were associated with the complaint being outbreak-associated. For individual characteristics we only examined information taken on the actual complaint caller. Variables assessed included number ill, age, diarrhea, fever, vomiting, bloody stool, health-care visit, and called a health-care provider. Chi-squares tests and t-tests were used to determine if there were significant differences between the groups. The Wilcoxon-Mann-Whitney test was used to test the difference between number of people in a complainant group that were ill and total number of exposed in a complainant group. A

logistic regression model was used to estimate associations between variables of interest from univariate analyses. Variables were included in the logistic regression model if they were found to be statistically significant in the univariate analysis at the $p < 0.05$ level. The outcome of this model was whether or not a complaint was outbreak-associated. Number exposed in party and number ill were highly correlated. We chose to include number ill in our multivariate model since it was likely to be more reliably answered. Number ill was categorized into one ill, two ill, or three or more ill.

We also examined a caller's ability to identify the source of their illness. This was done by examining the incubation period of illness of callers. Callers are asked about illness onset time and suspected meal time. Incubation periods were calculated by taking the difference in these times. Median incubation periods for outbreak and non-outbreak complaints were compared using a Wilcoxon-Mann-Whitney test.

3.3. Results

From 2000 to 2006, there were 332 confirmed foodborne outbreaks in the state of Minnesota; 261 (79%), were detected solely from consumer complaints. The majority of the complaints (162, or 62%) were about food establishments and 99 (38%) were about an event (Table 1). Of complaints about establishments, 12 (7%) were initially made to the establishment (and subsequently forwarded by the establishment to a health department) (Table 1). Eight (2%) additional outbreaks were detected through a combination of complaints and other surveillance methods. The remaining outbreaks

were detected through pathogen-specific surveillance, health-care provider reports of clusters of illness due to non-reportable pathogens, and reports from institutions such as schools and long-term care facilities (Table 1).

Sixty-one percent of outbreaks were caused by norovirus or suspected norovirus. The complaint system detected almost all of the outbreaks caused by non-reportable pathogens such as norovirus, *Clostridium perfringens*, *Bacillus cereus*, *Staphylococcus aureus*, and scombroid toxin (Table 1). Complaints also contributed to detection of outbreaks caused by reportable pathogens. For example, 5 (14%) *Salmonella* outbreaks were found solely through consumer complaints, and complaints contributed to outbreak identification in another 4 (11%) *Salmonella* outbreaks (Table 1).

From 2000 to 2006, the MDH foodborne illness complaint system received 5,414 complaints. The majority of these were received directly by MDH (Figure 1). We were able to identify complaints in the database associated with 236 of the 269 confirmed outbreaks that were found solely through complaints or with the aid of complaints. Of the 33 that we could not identify, 25 were reported to MDH from local health agencies and were not entered into the surveillance database. The remaining 8 outbreaks were found through direct consumer complaint calls, but no entry in the database could be identified. Some outbreaks were associated with multiple independent complaints, resulting in 359 (7%) complaints being designated as outbreak-associated. Of the 236 outbreaks identified in the complaint database, the majority, 165 (70%), resulted from one

complaint, with 146 (88%) of these reporting illness from multiple households with a clear common exposure. Outbreaks that resulted from one complaint reporting common exposure had a median number of 5 ill persons which was higher than outbreaks detected through multiple complaints except for one outbreak that resulted from 12 individual calls (Table 2).

The most predominant symptom of callers into the complaint system was diarrhea (83%). This was followed by cramps (77%), vomiting (66%), fever (25%), and bloody stool (4%). Among all callers to the complaint system, 870 (19%) had visited a health-care provider. Health-care visits were also stratified by the number ill in a complaint. The vast majority of calls only had one person ill. As the number of ill people reported in a complaint went up, the percent visiting a health-care provider went down: 22% for one ill, 15% for 2 ill, and 11% for 3 or more, $p < 0.001$. Although it was not possible to evaluate the frequency of physician visits by etiology, callers with both vomiting and diarrhea were more likely to visit physicians than were callers with other symptoms, 562 (25%) versus 308 (13%), respectively ($p < 0.001$). Examining lag times for complaints and reports obtained through pathogen-specific surveillance showed that lag times were much shorter for reports received through the complaint system, with a median time of 2 days versus 19 days ($p < 0.001$).

Complaint callers associated with an outbreak differed from non-outbreak complaint callers by several characteristics (Table 3). Outbreak-associated complainant callers were

slightly older, 45 versus 41 ($p=0.0002$) and were more likely to have diarrhea and fever than those from complaints that were not outbreak-associated. More non-outbreak-associated complainant callers also had sought medical care than those who were associated with an outbreak 834 [19%] vs. 36 [13%], respectively ($p<0.001$). The median number ill and the median number exposed were both greater for outbreak related complaints (Table 3). These variables were highly correlated.

The logistic regression model included age, number ill category, fever, diarrhea, bloody stool, physician visit, and physician call. Only age and number ill remained associated with outbreaks in the final model. The odds of a complaint being outbreak-associated were 16 times higher for complaints reporting 3 or more people ill than those with just one person ill in their group (Table 4). The odds of a complaint being outbreak-associated were also higher for complaints that had 2 people ill when compared to just one person ill.

Of the 5,414 complaints, 4,198 had complete time data to calculate an incubation period. We further excluded 109 complaints due to having illness onset times that preceded the meal time. The remaining 4,089 complaints included 264 outbreak-associated complaints. Median incubation times were 27 hours for outbreak-associated complaint callers, versus 6 hours for non-outbreak-associated complaint callers, $p<0.001$.

3.4 . Discussion

The aim of this paper was to examine the ability of a complaint-based surveillance system to detect foodborne illness outbreaks and to describe characteristics of a complaint-based surveillance system. Our analysis showed that consumer complaint systems are an effective surveillance tool for detection of foodborne illnesses caused by various etiologies, including reportable pathogens. Complaint systems have the ability to enhance pathogen-specific surveillance as well as provide the primary means of outbreak detection for non-reportable pathogens for which clinical laboratory diagnosis is not available.

The use of a complaint based surveillance system can also lead to faster detection of cases resulting in earlier investigations when compared to cases to be reported through pathogen-specific surveillance. Complaint systems have the ability to shorten the lag time between illness and report to the health department. This could lead to more timely investigations and follow up by health departments. This is especially important when items that are perishable, such as fresh produce, are the cause of outbreak. Also shorter lag times may lead to better food histories since recall is better closer to when the illness occurred. Improved information on exposures could lead to more reliable identification of outbreak sources.

Our data also indicated that a high priority should be placed on investigating calls reporting multiple illnesses. The number ill was significantly related to being an outbreak

associated complaint. This is related to the ability of the caller to determine a common exposure for their illness when more people are involved and the health departments targeting investigation of calls with clear common exposures. These types of calls represent events that are more likely to have been outbreaks and are readily detected through a complaint system that collects the proper information to allow for further follow up.

The majority of callers to the MDH complaint surveillance system never sought health care, so would never have entered the pathogen specific surveillance stream.

Interestingly, there appeared to be an inverse relationship between being an outbreak-associated complaint caller and seeking health care. Callers with more ill persons in their group were less likely to seek medical care. This could indicate that those who were part of a larger outbreak could self identify a cause better and thus did not seek medical care. It could also be due to norovirus causing larger numbers of ill individuals with less severe symptoms which would not need medical attention. Whatever the cause, it is clear that the vast majority of callers to complaint system would never enter the laboratory-based surveillance stream.

It has been previously shown that persons who attributed their illness to a meal eaten outside the home were likely to identify a meal eaten within 5 hours as the source of illness, even when such a short incubation period was inconsistent with their symptoms [41]. Our analysis found a similar result, with a median incubation period of 6 hours for

non-outbreak-associated complaints. However, for outbreak-associated complaints the median incubation period was 27 hours. A large proportion of outbreak-associated callers had a clear common source of exposure, such as a common event, or numerous ill from one meal with no other exposures, making identification of a source much easier. Those *without a common exposure, which accounted for the majority of non-outbreak-* associated complaints, tended to suspect the last meal eaten. It is clear when callers had enough information they were very adept at identifying a common exposure. When this was not the case callers were likely mistaken about the suspected cause of illness, supporting the need to collect detailed food histories.

This study had several limitations. There were several outbreaks found through complaints that we could not link to callers in the complaint database. Additionally, some callers may have been part of an outbreak that was never investigated by the health department. These limitations would tend to make differences between outbreak and non-outbreak related calls harder to detect. Even with these limitations we feel that this analysis shows the potential benefits of using consumer driven complaint systems for foodborne illness surveillance. This review of a robust complaint system illustrated the type of outbreaks detected, the response time, and source of identification. This has implications regarding potential interventions, especially for outbreaks involving food workers. Information on a caller's illness, establishments eaten at, and description of the number of people ill, are necessary to find food related outbreaks. Contact information of the caller is essential for further follow-up if an investigation is warranted. Complaint

systems fill many of the gaps in traditional laboratory surveillance. Using laboratory and complaint surveillance systems will give health departments needed tools to identify and prevent cases of foodborne illness. In addition to combining both surveillance methods it is also essential to act on gathered data. MDH's ability to detect a large number of foodborne outbreaks is due to their ability to aggregate data from various sources and aggressive follow up of potential cases.

3.5. Tables

Table 1: Source of case detection of confirmed foodborne outbreaks by etiology, Minnesota, 2000-2006

	Consumer Complaint		Provider Report (non-reportable disease)	Institution	Pathogen- Specific Surveillance			Multiple	Total
	Event	Food Establish- ment			Isolate submission	Provider report of lab- confirmed illness			
						Common Exposure	No common exposure		
Norovirus	62	91	2	8	0	0	0	1 ^b	164
Suspected Norovirus	19	17	1	1	0	0	0	0	38
<i>Salmonella</i> ^a	3	2	0	0	25	1	1	4 ^c	36
Diarrhea-Toxin Syndrome	5	12	0	1	0	0	0	0	18
<i>E. coli</i> O157:H7	0	0	0	0	8	1	2	3 ^d	14
Scombroid	0	10	2	0	0	0	0	0	12
Unknown	2	8	0	0	0	0	0	0	10
Vomiting-Toxin Syndrome	3	7	0	0	0	0	0	0	10
<i>Clostridium. Perfringens</i>	0	9	0	0	0	0	0	0	9
<i>Shigella</i>	1	0	0	1	3	0	0	0	5
<i>Bacillus. Cereus</i>	1	3	0	0	0	0	0	0	4
<i>Campylobacter</i>	1	1	0	0	1	0	0	0	3
Hepatitis A	0	0	0	0	0	2	0	0	2
Amatoxin	0	0	1	0	0	0	0	0	1
Astrovirus	1	0	0	0	0	0	0	0	1
<i>Cryptosporidium</i>	0	0	0	0	1	0	0	0	1
<i>Cyclospora</i>	0	1	0	0	0	0	0	0	1
Escolar gastroenteritis	1	0	0	0	0	0	0	0	1
<i>Listeria monocytogenes</i>	0	1	0	0	0	0	0	0	1
<i>Staphylococcus aureus</i>	0	0	1	0	0	0	0	0	1
Total	99	162	7	11	38	4	3	8	332

^a 1 outbreak attributed to consumer complaint event, had multiple pathogens positive in subsequent lab testing

^b 1 outbreak was detected through cases reported from an institution and consumer complaint about an event

^c 3 outbreaks had cases detection from complaints, and isolate submission, 1 outbreak found cases from complaints and provider report of confirmed case.

^d 2 outbreaks found cases from complaints and isolate submission, 1 found cases from isolate submission and provider report of a confirmed case

Table 2: Number of outbreaks in complaint database by the number of individual complaints associated with that outbreak in complaint system and median number ill per complaint

Individual Complaints	Number of Confirmed Outbreaks	Median Number Ill per Complaint (1Q, 3Q) ^a
1 ^b	143	5 (3, 12)
1 ^c	22	2 (1,3)
2	48	2 (1, 4)
3	12	2 (2, 3)
4	4	2 (2, 2.5)
5	3	3 (2, 3)
6	1	2 (1, 3)
7	2	2 (2, 3)
12	1	7 (5, 9)
Total	236	

^a Median number of ill individuals reported per complaint with 1st quartile and 3rd quartile

^b Individual complaint that was reporting illness from multiple households

^c Individual complaint reporting illness from single households

Table 3: Comparison of characteristics reported by the primary complaint caller by outbreak status

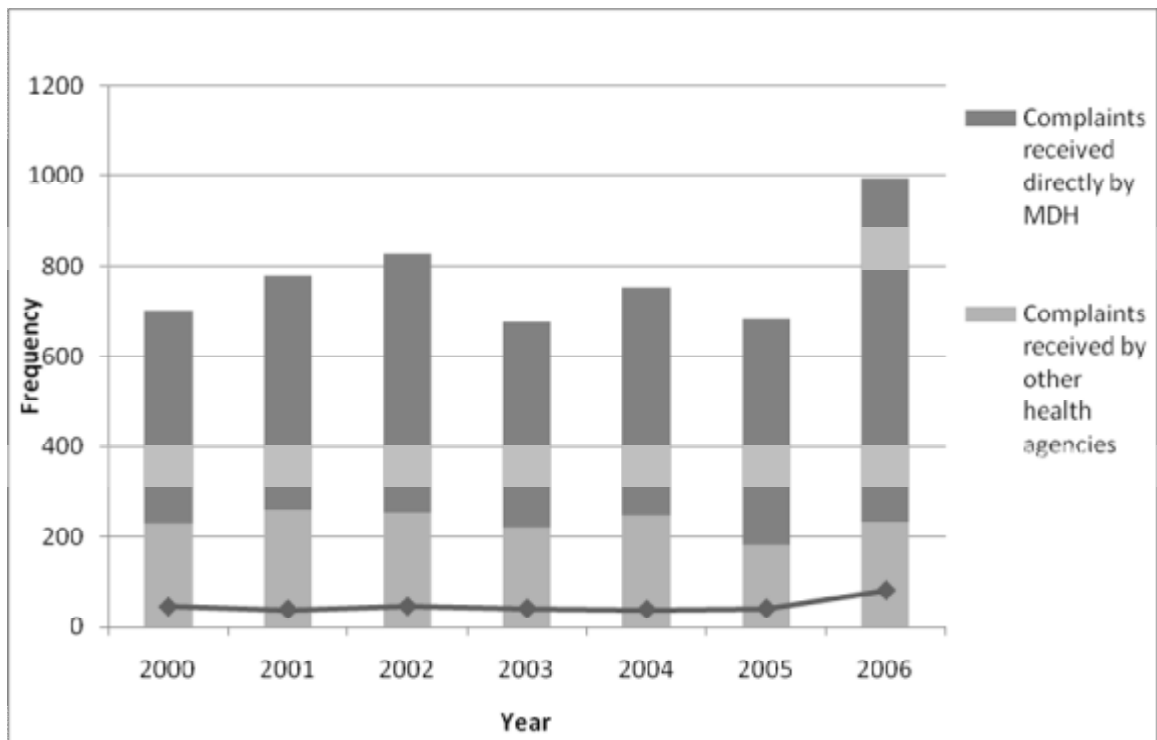
Variable	Associated with an Outbreak?		p-value	Missing (n=5414)
	Yes	No		
Age (years)	45	41	0.0002	848
Gender (% Female)	65	60	0.0812	96
Diarrhea (%)	89	85	0.021	96
Cramps (%)	81	78	0.25	394
Vomiting (%)	66	67	0.75	66
Fever (%)	43	36	0.042	1713
Bloody Stool (%)	3	6	0.016	1865
Called health- care provider (%)	21	31	0.0007	918
Visited health- care provider(%)	13	19	0.006	823
Number Ill in Party (median)	3	1	<0.001	447
Number Exposed in Party (median)	4	2	<0.001	492

Table 4: Results of final logistic regression model examining variables associated with being an outbreak associated call

Variable		Odds Ratio	95% Confidence Interval
Age		1.02	1.02-1.03
Number Ill	1	Ref	
	2	3.53	2.38-5.23
	≥ 3	15.8	11.1-22.5

3.6. Figures

Figure 1: Foodborne illness complaints by source of receipt and confirmed foodborne outbreaks by year, Minnesota, 2000-2006



Chapter 4. Manuscript 2

Title: Development of a *Salmonella* Screening Tool for Consumer Complaint Based Foodborne Illness Surveillance Systems

Foodborne illness surveillance based on consumer complaints detect outbreaks by finding common exposures among callers, but this process is often difficult. Lab testing of ill callers could also help identify potential outbreaks. However, collection of stool samples from all callers is not feasible. Methods to help screen calls for etiology are needed to more efficiently handle calls to complaint surveillance systems and increase the likelihood of detecting foodborne outbreaks caused by *Salmonella*. Data from the Minnesota Department of Health foodborne illness surveillance database were analyzed. Complaints with identified etiologies were examined to create a predictive model for *Salmonella*. Bootstrap methods were used to internally validate the model. Seventy-one percent of complaints in the foodborne illness database with known etiologies were due to norovirus. The predictive model had a good discriminatory ability to identify *Salmonella* calls. Three cutoffs for the predictive model were tested, one that maximized sensitivity, one that maximized specificity, and one that maximized predictive ability, providing sensitivities and specificities of 32% and 96%, 100% and 54%, and 89% and 72%, respectively. Development of a predictive model for *Salmonella* could help screen calls for etiology. The cutoff that provided the best predictive ability for *Salmonella* corresponded to a caller reporting diarrhea and fever with no vomiting, and 5 or less people ill. Screening calls for etiology would help identify complaints for further follow-up, and result in identifying *Salmonella* cases that would otherwise go unconfirmed; in turn, this could lead to the identification of more outbreaks.

4.1. Introduction

Foodborne illnesses are an important public health problem in the United States, causing an estimated 76 million cases annually at a cost of over \$30 billion [2, 4]. Preventing the distribution of contaminated food is difficult due to the complexity of the modern food supply and its myriad of potential points of contamination [74]. Early detection of outbreaks and removal of contaminated products is an important secondary prevention measure to limit the amount of illness [35].

The primary tool to detect outbreaks of foodborne illness is public health surveillance [36]. Public health surveillance for foodborne illnesses incorporates pathogen specific surveillance of reportable illnesses, and reports of illness through consumer complaints or clinicians. PulseNet, provides a platform for pathogen specific surveillance by allowing public health labs to share laboratory results to find clusters of illness [37, 39]. Work has also been done to apply cluster detection methods to PulseNet and other pathogen specific surveillance systems, including development of automatic detection algorithms for *Salmonella* cases [75]. These advances have helped detect numerous multi-state outbreaks due to *Salmonella* [76, 77]. However, many outbreaks likely go undetected or unsolved due to the small percentage of ill individuals who seek medical care and eventually have stool specimens submitted for testing. For every case of *Salmonella* reported through pathogen specific surveillance, it is estimated that 38 go unreported [4].

Consumer complaint based surveillance systems are also commonly used to identify foodborne illness outbreaks. Consumer complaint systems collect reports of foodborne

illnesses directly from the public, allowing individuals to report a suspected source of their illness. Typical information collected includes illness onset time and characteristics, the suspected food establishment, and a caller's food history. Callers are typically reporting food establishments, specific products, or events as the suspected source of their illness.

The majority of confirmed foodborne outbreaks are detected through complaint surveillance [45, 78]. Outbreaks are detected through complaint surveillance systems by one caller reporting illnesses in multiple households or multiple individual callers reporting illnesses associated with the same event or establishment. However, it is often difficult for the public to correctly identify the source of their illness [41]. Another drawback is that there is rarely information on the etiology of a caller's illness. Being able to determine the etiology of a caller's illness through laboratory testing would help in prioritizing which complaint should be investigated to identify potential foodborne outbreaks. However, testing of all callers is not feasible. A method to screen callers for etiologies of interest would increase the utility of a complaint based surveillance system.

Previous studies have examined the epidemiological profiles of foodborne outbreaks, including symptom profiles to determine etiologies of outbreaks. These studies used aggregate symptom information from individuals in an outbreak to determine possible causes of illness. *Salmonella*-like outbreaks were characterized by a high proportion of cases with fever and less with vomiting [49]. Norovirus outbreaks tend to have a higher proportion of cases reporting vomiting than fever [73].

Applying the principle of symptom profiling of outbreaks could provide a new method to identify the likely etiology of illnesses reported to complaint systems. To accomplish this, separate profiles are needed for each pathogen of interest. In this analysis we focused on *Salmonella*, the leading bacterial cause of foodborne illnesses [8]. The objective of this paper was to develop and validate a predictive profile for *Salmonella* based on data collected through a complaint surveillance system.

4.2. Methods

Data from the Minnesota Department of Health (MDH) foodborne illness surveillance database from 2000-2008 were used in this study. In our analysis we excluded data that were received through pathogen specific surveillance since we were primarily interested in direct consumer complaint surveillance. All callers to the system are asked a standard set of questions that include demographic and illness information for the caller, exposure information, and number of ill individuals associated with the suspected exposure [78].

Etiology of a complaint was determined in two ways: 1) if the caller had a laboratory confirmed infection, or 2) the caller was part of a confirmed foodborne outbreak.

Complainants that were suspected of being part of an outbreak were sent a stool collection kit to be returned to the MDH Public Health Laboratory for testing. If a complaint caller did not have a laboratory confirmed etiology but was associated with an outbreak with a known etiology this was used as the complainant's cause of illness. There were some confirmed outbreaks classified as being caused by bacterial intoxications where a laboratory confirmation was not available. In these cases the etiology was

classified as either a vomiting toxin or diarrhea toxin using the criteria described by Hall et al. [49]. All calls that were not identified as being associated with an etiology of *Salmonella* were categorized as non-*Salmonella* calls.

Univariate analysis was carried out on potential predictive variables for *Salmonella*. Variables examined were symptoms of a caller's illness, number ill reported by a caller, and age of a caller. These variables were chosen because they were individual level characteristics that are typically collected by complaint systems. Variables were tested using Chi-square tests, t-tests, or Wilcoxon-Mann Whitney test in cases where variables were not normally distributed.

Multivariable logistic regression was used to develop a predictive model for *Salmonella*. All variables from the univariate analysis that were significant at the 0.10 level were included in the multivariable model. The variables diarrhea, vomiting, fever, bloody stool, and the log of number ill were entered into the prediction model for *Salmonella*. The outcome of the model was if the etiology of the caller's illness was *Salmonella* or not. Variables that were not significant at the $p < 0.10$ level were removed from the model. Complaints that had missing covariate information were dropped from the multivariate analysis. The goodness-of-fit of the model was assessed with the Hosmer-Lemeshow test.

We differentiated between *Salmonella* calls and *Salmonella*-like calls. *Salmonella* calls were when the etiology of the caller's illness was known to be *Salmonella*. We used the

term *Salmonella*-like when referencing calls our model predicted were *Salmonella*, but the etiology was not known. The predictive ability of the model was examined using the area under the curve (AUC), which is identical to the c-statistic in logistic regression. The AUC measure quantified the ability of the model to discriminate between *Salmonella*-like calls and non-*Salmonella* calls. Bootstrap methods were used for internal validation, correcting for overfitting in our prediction model [66]. An optimism corrected AUC was calculated using the 0.632 method [60].

We also examined several thresholds for *Salmonella* prediction in the original data. The predictive model gives the probability that a caller's illness was due to *Salmonella*. For practical purposes it would be useful to convert this continuous probability into a binary prediction, *Salmonella*-like or not. To do this a threshold or cutoff for the probability is needed. We chose three cutoffs using a ROC curve for our model to accommodate the different goals health departments might have for their surveillance system. The first cutoff was chosen to maximize sensitivity, using the first threshold that reached 100% sensitivity. Another purpose of the tool could be to reduce unnecessary test by minimizing false positives. We set a maximum false positive rate of 5%. The second cutoff was the cutoff that fell below the maximum false positive rate. The third cutoff was chosen to maximize the discriminatory ability for *Salmonella* detection. In other words we chose the cutoff that made our binary prediction as close to perfect as possible. A perfect predictor has 100% sensitivity and 100% specificity and is represented by a single point on the ROC curve, the upper left corner of the graph. The distance from each threshold on the ROC curve to the upper left hand corner of the graph was determined

and the closest one was chosen as this threshold. Sensitivities, specificities, positive predictive values (ppv), and negative predictive values (npv) were calculated for each cutoff, with their 95% confidence intervals (CI).

Only about 10% of complaint calls from 2000-2008 had a known etiology and therefore could be used to create our predictive model. Once the model was created, the remaining 90% of calls without a known etiology were classified as *Salmonella*-like or non-*Salmonella*-like using the three cutoffs. These were totaled and multiplied by the ppv to estimate how many true *Salmonella* cases would have been detected using our predictive model during the time period. To examine seasonal trends we also plotted monthly counts of *Salmonella*-like calls versus non-*Salmonella*-like calls as determined by our cutoff that provided the maximum discrimination.

4.3. Results

There were 7,059 complaints in the MDH foodborne illness complaint surveillance database during the study period. An etiology was determined for 683 of the complaints. There were 538 (79%) complaints with a laboratory-confirmed etiology in the database. We determined the etiology of the remaining 145 (21%) complaints by examining the etiology of the outbreak to which the caller was associated. Of the total complaint calls with known etiologies, 483 (71%) were due to norovirus. The second leading etiologies causes were *Salmonella* and *Clostridium perfringens*, with 37 complaint calls each (Table 1).

In the univariate analysis, selected caller's symptoms and the number ill they reported were associated with *Salmonella* as the etiology (Table 2). The number ill in a complaint was left skewed, with the majority of calls only reporting one ill. The median number ill in a call was found to be significantly different between *Salmonella* and non-*Salmonella* complaints, with a median of 1 versus 3 (Table 2). The age of the caller was not a predictor for *Salmonella*.

The final model contained the covariates of fever, vomiting, and log of number ill (Table 3). The Hosmer-Lemeshow test was not significant ($p=0.98$), indicating no evidence of lack of model fit. The AUC was 0.89 (95% CI, 84-93) with an adjusted AUC of 0.88. A model was also run including *Campylobacter* and *Shigella* calls with *Salmonella* as cases. This model had similar results to the original model that only had *Salmonella* as cases in the outcome, with vomiting, fever, and log of number ill remaining the only significant predictors (data not shown).

The cutoffs used to examine sensitivity and specificity were chosen with the aid of the ROC curve (Figure 1). This first cutoff that was most sensitive for *Salmonella* had a sensitivity of 100% (95% CI, 85-100), specificity of 54% (95% CI, 50-58) and a ppv of 10% (95% CI, 7-14), and corresponded to a caller reporting vomiting, fever, and three or fewer individuals reported in the call. Using the cutoff that was most specific for *Salmonella* gave a sensitivity of 32% (95% CI, 17-52), specificity of 96% (95% CI, 94-97) and a ppv of 28% (95% CI, 14-47). Callers fitting this cutoff had similar symptoms to that of the first cutoff but with only one ill individual reported. The third cutoff that

maximized the predictive ability for *Salmonella* gave a sensitivity of 89% (95% CI, 71-97), specificity of 72% (95% CI, 68-75), and a ppv of 13% (95% CI, 9-19) (Table 4). This cutoff corresponded with a caller having fever with no vomiting, and five or fewer ill reported in the call. Diarrhea was not included in the final model or cutoffs since almost all callers experienced diarrhea including every caller with *Salmonella*. From 2000-2008, the complaint surveillance system received a 7,059 complaints. Out of the total number of complaints, 2,886 matched the cutoff providing the best predictive ability for *Salmonella*. Applying the positive predictive value of 13% to these *Salmonella*-like calls resulted in 375 potential cases over the 9-year period (Table 4).

Figure 2 shows the frequency of complaint calls by month for *Salmonella* like calls and non-*Salmonella*-like calls as determined by our cutoff providing the best discriminatory ability. The non-*Salmonella* like calls show a rise through the winter months, which parallels the seasonality seen with norovirus outbreaks. *Salmonella*-like calls remain relatively stable throughout the year.

4.4. Discussion

The aim of this study was to develop a predictive model for *Salmonella* based on consumer complaint surveillance data. The use of clinical profiles has been useful to help determine the etiology of an outbreak. We extended this idea to predicting an etiology of a complainant's illness. Our predictive model was able to identify *Salmonella*-like calls, allowing health departments to focus investigations on calls that have an increased likelihood of being caused by *Salmonella*.

We presented 3 cutoffs that could be used in *Salmonella* prediction. Each has its drawbacks, but they illustrate how a predictive model could be tuned to the needs of an individual health department. The cutoff that maximizes predictive ability might provide the most value. For our 9-year study period 2,886 complaints met this cutoff, or 321 a year. The positive predictive value indicated that 375 would be potential cases, resulting in an estimated 42 additional *Salmonella* cases a year. For comparison in 2007, there were 709 culture-confirmed cases of *Salmonella* infection in Minnesota [79]; an additional 42 cases would have represented a 6% increase in *Salmonella* cases found at a cost of 321 additional tests.

The positive predictive values for all cutoffs were low, ranging from 10% to 28%. However, this may not be as much of an issue as it is in traditional screening tests. The purpose of this screening is not to determine an exact diagnosis, but to identify calls that fit a profile that that could be targeted for follow up. The marked difference in seasonal trends when we compared monthly counts of *Salmonella*-like and non-*Salmonella* complaint calls helps illustrate the ability of our predictive model in differentiating etiologies. Norovirus has a strong seasonal component with more cases in winter [80]; this seasonality was not seen when we plotted monthly counts of *Salmonella*-like calls. This provides additional evidence that we were able to separate out norovirus calls from *Salmonella*-like calls.

A limitation of this study was the low number of calls with a known etiology. The majority of callers to the complaint system are never tested to determine the cause of their illness. Most of those tested were either part of an outbreak or suspected to have been part of an outbreak. This has the potential to bias our findings if *Salmonella* cases were more likely to have been tested due to risk factors used in our prediction model. However, this is not likely. The health department did not initiate investigations based on symptoms of callers. Investigations were usually initiated when a common exposure could be clearly identified. The etiology of a caller's illness was seldom known prior to initiation of the investigation. Another limitation was the low number of *Salmonella* cases represented by callers. There were likely many more *Salmonella* cases that were never identified. These cases that were never investigated may not fit the profile that was created for known *Salmonella* cases. In this analysis we also aggregated all non-*Salmonella* cases together. Norovirus made up the majority of these non-*Salmonella* cases, but there were also other pathogens such as *Shigella* and *Campylobacter* that cause similar symptoms as *Salmonella*. We also ran our analysis with these pathogens included in the outcome with *Salmonella* cases. Results from this were essentially the same indicating that our prediction model for *Salmonella* would also identify possible *Shigella* and *Campylobacter* calls when used to screen complaints.

Complaint surveillance systems are an effective way to detect foodborne outbreaks. Developing and using predictive profiles are ways to help add to the utility of these systems. Underreporting of foodborne illnesses is a large problem. The ability to find a source of illness is dependent on having enough cases to triangulate a common exposure.

Narrowing down individuals for follow up and lab testing could lead to more information to link cases in the complaint database as well as the ability to link to pathogen specific surveillance. To further research in this area more testing should be done on causes of illness in callers to complaint systems. Information on sporadic cases could help in development of better predictive models and enhance case identification. Exploring the use of predictive models and application of new methods to analyze incoming calls at the local and state level could greatly help in targeting follow-up of potential foodborne illness leading to detection of more cases on interest and increasing the likelihood of identifying a common source of illness.

4.5. Tables

Table 1. Complaints from MDH complaint surveillance database from 2000-2008 with identified etiologies

Etiology	Number of Complaints (%) (n=683)
Norovirus	483 (71)
<i>Salmonella</i>	37 (5)
<i>Clostridium perfringens</i>	37 (5)
Diarrhea Toxin Syndrome	27 (4)
<i>Campylobacter jejuni</i>	19 (3)
Vomiting Toxin Syndrome	13 (2)
Scombroid	8 (1)
<i>Staphylococcus aureus</i>	5 (0.7)
<i>Shigella spp.</i>	4 (0.6)
<i>Bacillus cereus</i>	4 (0.6)
<i>E. coli</i> O157:H7	3 (0.4)
Other	43 (6)

Table 2. Univariate analysis of potential predictive variables to distinguish *Salmonella* calls from non-*Salmonella* calls

Variable	<i>Salmonella</i>	Non- <i>Salmonella</i>	p-value
Diarrhea (%)	37 (100)	555 (86)	0.038 ^a
Vomiting (%)	11 (30)	403 (62)	0.003
Fever (%)	26 (70)	197 (31)	<0.001
Bloody Stool (%)	8 (22)	24 (4)	<0.001
Cramps (%)	33 (89)	469 (73)	0.142
Age (SD)	37 (13)	42 (18)	0.321
Number Ill (q1, q3) ^b	1 (1,2)	3 (2, 6)	0.003 ^c

^a Fishers exact test was used due to small numbers in some cells

^b Median with first and third quartile

^c Wilcoxon-Mann Whitney Test

Table 3. Results of multivariate logistic regression to examine predictive variables a complaint being a *Salmonella* case

Variable	OR (95% CI)	
	Full Model	Final Model
Diarrhea	>99 (<.01, >99)	-
Vomiting	0.22 (0.09, 0.53)	0.22 (0.09,0.53)
Fever	3.22 (1.28, 8.06)	3.89 (1.64, 9.20)
Bloody Stool	1.59 (0.44, 5.70)	-
Log Number Ill	0.18 (0.07, 0.44)	0.17 (0.07, 0.42)

Table 4. Sensitivity, specificity, positive predictive value, and negative predictive value for 3 separate cutoffs for *Salmonella*, maximizing both specificity and sensitivity, maximizing specificity, and maximizing sensitivity

Cutoff	Sensitivity (95% CI)	Specificity (95% CI)	PPV	NPV	<i>Salmonella</i> - like calls 2000-2008 ^a	Potential cases ^b
Max sensitivity	100 (85, 100)	54 (50, 58)	10 (7, 14)	100 (98, 100)	5,030	503
Max specificity	32 (17, 52)	96 (94, 97)	28 (14, 47)	97 (95, 98)	242	68
Max prediction	89 (71, 97)	72 (68, 75)	13 (9, 19)	99 (98, 100)	2,886	375

^a Complaints in Minnesota from 2000-2008 that met the cutoff for *Salmonella* prediction

^b Potential cases calculated by multiplying the number of *Salmonella*-like calls by the PPV

4.6. Figures

Figure 1: ROC curve of predictive model trained on original data

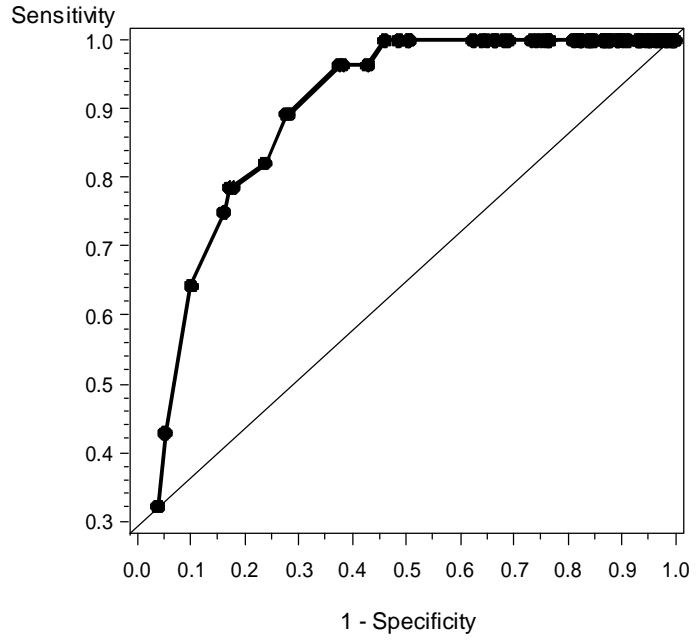
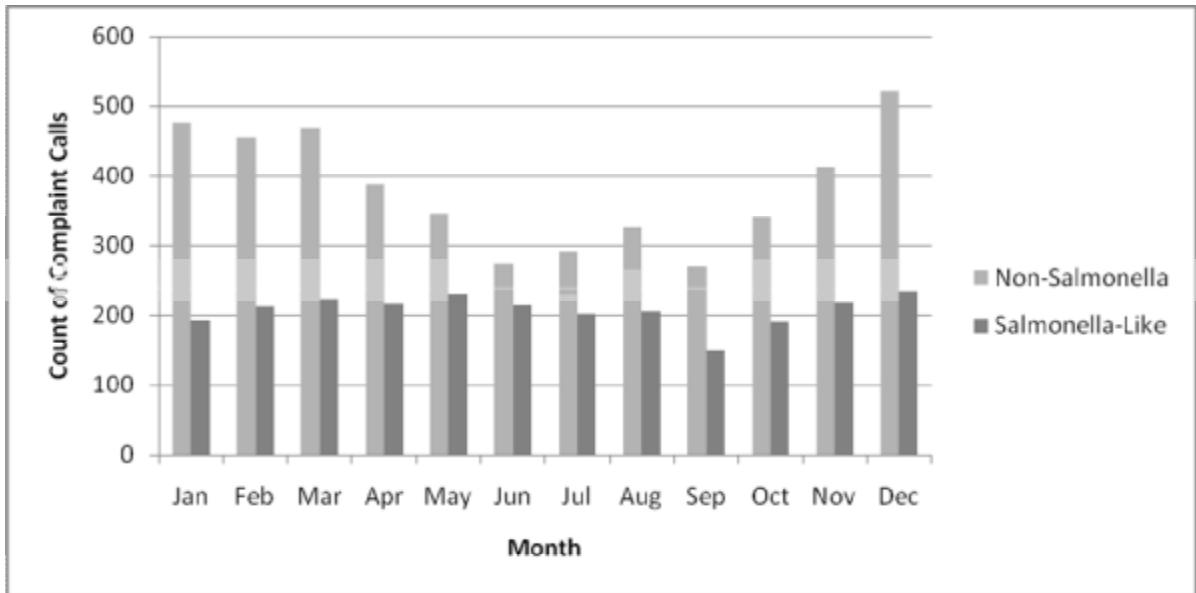


Figure 2: Count of complaint calls by *Salmonella* prediction by month



Chapter 5. Manuscript 3

Title: An Algorithm for Detection of Foodborne Outbreaks in Consumer Complaint Based Surveillance

Foodborne illness outbreak detection is becoming more difficult due to changing eating habits and the myriad of products available to the public. Complaint based surveillance systems provide a method to detect foodborne outbreaks from a variety of causes. However, methods to prospectively examine incoming data have not been well developed.

Cumulative sum methods have been applied to influenza surveillance and laboratory-based *Salmonella* surveillance. These methods provide a way to flag unusually high occurrences of events in surveillance data. This paper developed a cumulative sum algorithm for complaint based surveillance data to detect possible outbreaks. Weekly complaint calls in Hennepin County, Minnesota from 2000-2006 were adjusted for season and autocorrelation. Mean residuals and standard deviations from this model were used to calibrate a cumulative sum algorithm. The cusum algorithm was applied to 2007 data to evaluate its ability to detect outbreaks.

Two detection levels were used for our algorithm. The lower threshold (0.5 standard deviation above the mean) flagged weeks corresponding to 6 of 17 outbreaks that were identified in Hennepin County during 2007. The six outbreaks included 5 norovirus outbreaks and one large multi-state *Salmonella* outbreak. The algorithm flagged weeks corresponding to all outbreaks that were detected through multiple individual complaints. Sensitivity and specificity for the detection algorithm were 26% and 81% at the lower threshold when looking at all outbreaks. When focusing only on norovirus outbreaks the sensitivity and specificity were 63% and 84%.

Use of this outbreak detection method may help identify outbreaks with multiple individual complaints, and may be more applicable to norovirus outbreaks or those that result in large numbers of cases. It would be particularly useful to health departments that do not have a mechanism to regularly review incoming data for common exposures or do not collect enough information to make this possible. Cusum algorithms provide an easy and low cost method to automatically monitor incoming surveillance data to flag unusually high numbers of complaint calls in a given time period.

5.1. Introduction

Primary prevention of foodborne illnesses is becoming more difficult due to the increasing complexity of the food supply system [7]. Secondary prevention methods for reducing the burden of disease rely on early detection of foodborne disease outbreaks and removal of contaminated product [35]. This relies on identifying ill individuals and linking them to a common exposure. However, this is a complicated task due to issues of underreporting, inaccurate identification of suspected cause of illness, and lack of resources needed to fully investigate each suspected case. Additionally, being able to successfully identify a common exposure often depends on having a large number of ill individuals to interview.

The primary tool to find cases foodborne illness in the general population is public health surveillance [36]. Finding outbreaks through surveillance requires the public health practitioner to find commonalities among ill individuals. Surveillance for foodborne illnesses is typically based on either pathogen-specific surveillance of reportable illnesses, reports of clusters of illness by astute clinicians, or complaint reports by the general public. Each method has its drawback, but when used in concert provides a comprehensive means to find cases of foodborne illness in the community from a wide range of etiologies [44, 78].

The definition of a foodborne outbreak is when two or more persons experience similar illnesses after the ingestion of a common food item. This is often operationalized as two or more persons in separate households. Outbreaks are traditionally detected in complaint

based surveillance by callers identifying a common source of exposure among multiple ill individuals in their party, or linking multiple individual calls to a single exposure.

However, the general public often doesn't recognize the correct cause of their illness, generally attributing their illness to the last thing eaten when incubation times for most foodborne illnesses are typically much longer [41]. Follow up investigations by health departments are crucial to find outbreaks in all surveillance systems including complaint based surveillance, but in many instances it is not clear if follow up is needed. Outbreaks that are caused by a widespread exposure would not show an obvious link between cases. Methods to detect clusters of illness by examining changing patterns in call volume are needed to increase the utility of complaint based surveillance systems. Determining unusually high call volume would indicate a possible occurrence of an outbreak that would trigger further investigation of callers.

Cumulative sum methods (cusum) have been suggested as a method to detect rises in illness in syndromic surveillance data [81, 82]. Cusum methods have the ability to detect small shifts in the mean quickly, allowing their use in a prospective manner on incoming data. These methods, while traditionally used in industrial processes, have been applied to several health examples. Cusum based detection algorithms monitoring *Salmonella* lab data have been responsible for detection of several multi-state outbreaks [72, 75]. The utility of cusum methods have also been demonstrated in influenza surveillance, looking for unusual amounts of reported influenza-like illness in hospital discharge data [83]. This same methodology can be applied to foodborne illness complaint surveillance, looking for unusual rises in complaint calls, alerting health officials for the need to follow

up on clusters of illness that may not appear related. Additionally, cusum detection algorithms could provide an automated method to monitor incoming surveillance data for agencies that do not have a mechanism in place to review data for common exposures or do not collect enough detailed information to identify common exposures.

Consumer complaint surveillance has the ability to detect foodborne illness outbreaks, but few prospective methods to analyze and more effectively use incoming data have been developed. The objective of this paper was to develop a cusum algorithm to detect outbreaks based on increased volume of calls and applying it to complaint data from a specific geographic area in 2007.

5.2. Methods

Data from the Minnesota Department of Health (MDH) foodborne illness complaint surveillance database from 2000-2007 was used in this analysis. Information in the database comes primarily from direct complaints of foodborne illness from the general public. Complaints can also come directly from food establishments who have received complaints from the public. Notification of illness clusters can also come from institutions such as schools and nursing homes, and from clinicians. The complaint surveillance system is run by MDH and provides statewide coverage. Local health departments can also complete a standard complaint questionnaire and forward the information to MDH. The complaint questionnaire collects contact information for caller, illness information, suspected establishment, number ill and exposed, and a 4-day food history when no common exposure in multiple households is indicated. For this analysis

we focused complaints from Hennepin County residents. Hennepin County is the most populous county in Minnesota (1,141,000 people) and contains the city of Minneapolis.

Complaints were coded by the week of the year they were received. Each year was split into 0-53 weeks, with a week starting on a Sunday. Weeks 0 and 53 could have been partial weeks. Complaints were summed by the week of the year they were received and were also coded by the season they were received: fall, winter, spring, or summer. Each season contained 3 months, with September, October and November making up fall; December, January, and February as winter; March, April and May as spring; and June, July and August as summer.

Cusum methods were used to develop an algorithm to detect unusual numbers of complaint calls. Basic cusum methods require two assumptions to be satisfied: 1) the value being monitored is distributed normally, and 2) the value is not autocorrelated [81]. These are often violated with health data. Health data are frequently autocorrelated, since they are often time series data in which observations close in time are correlated. However, remedial measures to address these assumptions have been suggested. Although residuals do not always have a normal distribution, they have been shown to work well in cusum tests [84]. An autoregressive model, using proc autoreg in SAS v.9.2, was used to model the weekly counts of complaint calls in Hennepin County during 2000-20006. The model closely resembles an ordinary linear regression model with the addition of an autoregressive error term. We examined if the distribution of the residuals and count of complaints were normal by plotting histograms for both. Additionally we

checked to make sure that there were no negative predictive values from our model. Data from complaint calls is not likely to be independent since counts of complaint calls are correlated between weeks. Additionally, data in adjacent weeks are correlated with each other. Using an autoregressive model adjusting for this autocorrelation allows the error term to be independently distributed, allowing the use of residuals for the cusum algorithm.

Complaints were summed by the week of the year they were received. Norovirus illness, which account for the large majority of calls, have a strong seasonal trend with more cases in the winter months [17]. Due to this, we adjusted for season in our model by including indicator variable for spring, summer, and fall, with winter absent as the reference season and the weekly count of calls as the outcome. The model also included a second order autoregressive error term that was selected by stepwise autoregression. The stepwise autoregression first started with a 5th order error term and removed terms sequentially until all autoregressive parameters were significant at the $p < 0.05$ level. Residuals from the model were then analyzed using cusum methods.

In order to use cusum methods, a delta, or a shift in the process mean that our cusum is designed to detect needed to be determined. In our analysis, delta was the change in the size of the residual that we want to detect. A one standard deviation change is often set as delta. However, it is also common practice to set delta to some meaningful value for the process under study [85]. In the case of complaints data it is not clear what a meaningful value for a change in the mean residual would be. For this reason we chose two levels of

delta, 0.5 and 1 standard deviation, to examine the differences in outbreak detection. A one-sided cumulative sum was used in this analysis, since we were primarily interested in a rise in complaint calls, indicating a possible outbreak. An one-sided cusum can focus on an upper or lower limit. In this case the cusum only detects changes that are higher than our set detection level.

The cusum algorithm tuned to mean historical data from 2000-2006 was applied to 2007 data with detection levels set to 0.5 and 1 standard deviation. The cusum indicated weeks where the residuals of the autoregressive model adjusting for seasonality were more than expected; meaning the number of callers was higher than expected for that week. We referred to these as flagged weeks.

Flagged weeks were compared to confirmed foodborne outbreaks in Hennepin County in 2007. We categorized all outbreaks by etiology and detection method. Detection method refers to the manner in which MDH found out about the outbreak. This could have been through one of several possibilities: pathogen-specific surveillance, clinician report, individual complaint from the public about an establishment or event, and multiple independent complaints from the public about an establishment. The week of the year that MDH was first alerted of the outbreak was determined. We also determined the week of illness onset. This could have spanned multiple weeks in the case of a propagated outbreak with cases over multiple weeks.

We examined the sensitivity, specificity, positive predictive value (PPV), and negative predictive value (NPV) of the algorithm by comparing flagged weeks to outbreak weeks. Outbreak weeks were any weeks where cases from an outbreak had occurred. Outbreaks were found through a variety of methods. It is unlikely that our detection algorithm would detect small outbreaks found through pathogen-specific surveillance. Due to this we calculated sensitivities, specificities, PPV, and NPV for all outbreaks as well as for norovirus outbreaks in particular, since there is no pathogen-specific surveillance to pick up norovirus outbreaks. Sensitivity was calculated as the number of flagged weeks that were associated to outbreaks divided by the total number of outbreak weeks. Specificity was calculated as the number of weeks that were not flagged divided by the total of weeks that did not contain outbreaks. PPV was calculated as the number of flagged weeks that corresponded to outbreak weeks divided by the total number of flagged weeks. NPV was the number of unflagged weeks that were not associated to outbreak divided by the total number of unflagged weeks.

5.3. Results

From 2000-2006 there were a total of 5,414 complaints made to MDH; county of residence was noted for 4,701 (87%) complaints. Of the total complaints, 1,538 (28%) were from Hennepin County residents. The weekly mean of complaints for this county was 4.3 with a standard deviation of 2.5. The autoregressive model showed that there was a significant seasonal effect. Calls were highest in the reference season, winter. Examining the parameter estimates for the dummy variables for season indicated that there were approximately 1.2 fewer calls per week during the spring, and 2 less calls per

in summer and fall when compared to winter (Table 1). The mean residual for the autoregressive model using 2000-2006 data was 0.001 with a standard deviation of 2.3.

In 2007, there were 17 confirmed foodborne outbreaks in Hennepin County, that corresponded to 27 weeks where illnesses were detected. There were 5 outbreaks where illness onset for cases were spread over several weeks. Eight of these were caused by norovirus (Table 2). Nine of the outbreaks were found directly through consumer complaints; the others were detected through pathogen-specific surveillance or reports from clinicians. Weeks of the year were numbered 0-52 with weeks 0 and 52 being partial weeks. Week 0 had 6 days and week 52 had 2 days. During this same period, 144 complaint calls were received from residents of Hennepin County.

The cusum algorithm with a detectable difference of 0.5 standard deviations resulted in 12 flagged weeks; 7 of these were weeks in which outbreaks occurred (Table 2). The algorithm flagged all three outbreaks that resulted from multiple independent complaints of the same establishment. Flagged weeks corresponded with 1 *Salmonella* outbreak and 5 norovirus outbreaks. The *Salmonella* outbreak, caused by commercially distributed leafy greens, was originally detected through pathogen-specific surveillance. No cases were detected through complaints data. However, the outbreak was fairly large and covered a 2-month time period with cases found in multiple states [86]. The sensitivity and specificity for detection of all outbreaks was 26% and 81% respectively, with a ppv and npv of 58% and 51%. For norovirus outbreaks only, the sensitivity was 63%, specificity was 84%, ppv was 42%, and npv was 93% (Table 3).

When the detectable difference was 1 standard deviation there were 7 flagged weeks, with 3 weeks corresponding to outbreaks. One included the multi-state *Salmonella* outbreak and 3 included norovirus outbreaks (Table 2). The sensitivity and specificity for detection for all outbreaks was 11% and 77% respectively, and the PPV and NPV were 33% and 45%. For norovirus outbreaks only, the sensitivity and specificity were 25% and 87%, and the ppv and npv were 22% and 87% (Table 3).

5.4. Discussion

The purpose of this analysis was to develop a cusum algorithm for outbreak detection using foodborne illness complaint-based surveillance data and to assess its ability to detect outbreaks in real complaint data. Cumulative sum methods have been applied to pathogen-specific surveillance to detect dispersed outbreaks that would otherwise go undetected. These same methods could be used to automatically flag incoming complaint surveillance data for weeks with unusually high call volume.

This method may be especially useful in identifying outbreaks involving multiple independent complaints about the same establishment. In our analysis, the cusum algorithm flagged weeks corresponding to every confirmed outbreak detected through multiple independent complaints. This could simply be because bigger outbreaks tend to result in more complaints. Health departments that continually review incoming data already regularly identify this type of outbreak. However, if surveillance systems grow to capture more data, multiple independent complaints that mention the same establishment

might be harder to link due to more incoming data. Additionally, it is often the case that common establishments among multiple independent complainants are identified only when data on multiple potential exposures are collected (e.g., through a detailed 4-day food history). For jurisdictions that do not take a detailed food history, this would be difficult. In this case automatic methods to detect possible clusters of calls would be also beneficial. Weeks with outbreaks found through individual complaints that reported illness in multiple households that had a common exposure, or through pathogen-specific surveillance, were not as readily flagged using cusum methods. However, health departments may already be more apt to detect these types of outbreaks.

In our analysis, both detection levels also flagged a week during a *Salmonella* outbreak found through pathogen-specific surveillance. This was a multi-state outbreak that resulted in a relatively large number of cases. It is plausible that ill individuals may have reported to the complaint hotline during this period without being able to pinpoint their source of illness. It is possible that investigation of these illnesses may have detected more cases related to the outbreak.

Our measures for sensitivity and specificity showed that a lower detection threshold is better at detecting outbreaks. The 0.5 standard deviation threshold provided a sensitivity of 26% and a specificity of 81% for all outbreaks, and a sensitivity of 63% and a specificity of 84% for norovirus outbreaks. Positive predictive values were 58% and 42% for all outbreaks and norovirus only outbreaks, respectively. In other words, a flagged week corresponded to a week with a confirmed outbreak 58% of the time, and to a week

with a confirmed norovirus outbreak 42% of the time. This may be low in comparison with other diagnostic tests, but as a mechanism for narrowing down complaints to investigate would still be useful. Cusum algorithms would not detect all types of outbreaks, but by indicating calls that otherwise might not be investigated it would help identify outbreaks that might otherwise be missed. The negative predictive value for norovirus was quite high at 93%, meaning that for unflagged weeks there was a 93% chance that there really was no norovirus outbreak that led to being a confirmed outbreak.

Using these measures is a good starting point to evaluate the ability of the algorithm to find outbreaks, but it does have its limitations. As our gold standard measurement we used confirmed foodborne outbreaks as identified by the state health department.

However, this is not a true measure of all the outbreaks that occurred. Many outbreaks were likely unidentified due to an inability to find a common cause, either because too few cases were reported or an investigation was never initiated. It is unclear if our false positives were truly false positives or outbreaks that were never identified. A previous study of complaint based surveillance indicated that there were likely a larger number of outbreaks in the data that were never investigated [45]. In a public health setting PPV would likely be the measure of most interest, since it tells you how likely a flagged week would be the result of an outbreak. In this analysis, we provided PPV estimates for our algorithm. However, more research would need to be done to find more accurate PPVs since some flagged weeks could have been true outbreaks. Additionally, we used outbreak weeks as our denominator so one outbreak could have spanned multiple weeks. The cusum algorithm often did not flag all weeks related to an outbreak. If we look just at

outbreaks detected with the lower threshold (as opposed to all weeks involved in outbreaks), the algorithm flagged 8 of 12 (67%) of all outbreaks, and 5 of 8 (63%) of norovirus outbreaks.

The cusum algorithm for both levels of detection flagged multiple weeks of unusual calls in December. This corresponded to several norovirus outbreaks in the system. However, there were some flagged weeks where a confirmed outbreak did not occur. Even after adjusting for seasonality it appeared that December 2007 had an unusually high call volume. In this case it appears that the cusum algorithm could have been detecting a period of particularly high norovirus activity, in essence working as a syndromic surveillance system. Syndromic surveillance has previously been used to identify clusters of illness in surveillance datasets [87, 88]. However, this type of surveillance often suffers from low specificity [51]. Using complaint surveillance data would help address some of the issues of specificity since callers already self identified as having a foodborne illness.

There were several limitations to this study. We were not able to identify the county of residence for all calls to the complaint system. This could have biased our results if identification of county was related to if a call was truly an outbreak. However, there is no reason to believe that those that refuse to identify their county of residence were more or less likely to have been part of an outbreak. Another limitation is the uncertainty on how to geographically locate complaints. Due to the reliability of the data and how investigations are typically carried out we chose to analyze data on the county of

residence instead of the county where the implicated establishment was. Using location of the suspected establishment may yield a better alarm system; however, this information is not always as readily collected and many people implicate the wrong establishment.

This analysis has shown the possibility of using a cusum algorithm to detect outbreaks at the county level. The algorithm may complement current surveillance practices even where many complaints are already investigated. MDH always initiates outbreak investigations when multiple individual callers to the complaint system report the same suspected source of illness. The difficulty is that callers are not always able to correctly identify their source of illness. Our algorithm had several false positives which could have been potential outbreaks. For health departments that do not have an existing system to link multiple individual calls, the cusum algorithm could provide an easy method to identify groups of calls to investigate further. Additionally, complaint based surveillance systems and cusum based automatic detection algorithms may be able to act as an early warning system for increased illness activity. This would not only alert health departments for the need to increase investigations, but also allow them to send out warnings to local food establishments, nursing homes, schools, and other institutions particularly vulnerable to foodborne outbreaks. As public health surveillance systems for foodborne illnesses grow and become more complicated automated detection methods such as this will need to be applied to efficiently utilize incoming data.

5.5. Tables

Table 1: Regression coefficients from autoregressive model of complaint data from Hennepin County, 2000-2006

Variable	Parameter Estimate	95% Confidence Interval
Spring	-1.13	-1.55, -0.30
Summer	-1.97	-2.81, -1.12
Fall	-1.95	-2.78, -1.12
Winter	Ref.	
Intercept	5.63	

Table 2: Outbreaks in Hennepin County in 2007 and flagged weeks from the cusum detection algorithm

Month	Week of Report to Health Department	Week(s) of illnesses	Etiology	Detection Method	Flagged Weeks	
					$\Delta = 0.5$	$\Delta = 1$
January	1	1	Norovirus	Individual complaint ^b		
January	4	3	Norovirus	Individual complaint		
January	7	6,7	Hepatitis A virus	Clinician report		
Mar-April ^a	14	11-14	<i>Salmonella</i>	Pathogen-specific surveillance	12-14	13
April	16	16	Norovirus	Multiple complaints ^c	16, 18	
June	23	23	Norovirus	Individual complaint		
June	25	25	Scombroid Toxin	Report of complaints from restaurant		
July	29	28	Suspected Bacterial Toxin	Individual complaint		
May-July	29	22-26	<i>Salmonella</i>	Pathogen-specific surveillance		
July	29	28	Scombroid Toxin	Clinician report		
August	33	30-31	<i>Salmonella</i>	Pathogen-specific surveillance		
August	32	31	<i>Vibrio parahaemolyticus</i>	Pathogen-specific surveillance		
September	37	37	Scombroid Toxin	Clinician report		
November	47	45	Norovirus	Multiple complaints	45, 47	45, 47
December	49	49	Norovirus	Individual complaint	48-52	47-51
December	49	48,49	Norovirus	Individual complaint	48-52	47-51
December	52	52	Norovirus	Multiple complaints	48-52	47-51

^a large multiple-jurisdiction outbreak, ultimately resulting in 76 identified cases in 14 states

^b one complaint led to detection of this outbreak, complaint could have been reporting on

multiple ill individuals with a common source of illness

^c multiple complaints were received implicating a common source of illness

Table 3: Sensitivity, specificity, positive predictive value, and negative predictive value for outbreak detection using the cusum algorithm at two detection levels

	Detectable Difference ^a	Sensitivity ^b	Specificity ^c	Positive Predictive Value	Negative Predictive Values
All Outbreaks	0.5	7/27 (26)	21/26 (81)	7/12 (58)	21/41 (51)
	1	3/27 (11)	20/26(87)	3/9 (33)	20/44 (45)
Norovirus Outbreaks Only	0.5	5/8 (63)	38/45 (84)	5/12 (42)	38/41 (93)
	1	2/8 (25)	39/45 (87)	2/9 (22)	39/45 (87)

^a Detectable difference set for cusum algorithm, 0.5 and 1 standard deviations

^b Number of flagged weeks that corresponded to an outbreak divided by number of outbreak weeks

^c Number of non-flagged weeks divided by number of non-outbreak weeks

Chapter 6. Manuscript 4

Title: Complaint-Based Surveillance for Foodborne Illness in the United States: A Survey of Local Health Departments

Foodborne illnesses are an important public health problem, both in terms of the burden of illness and cost to the health care system in the United States. Preventing outbreaks is becoming more difficult due to the increased complexity of the food supply chain. Strengthening foodborne illness surveillance could help address the growing issues of food safety in the United States. Although public health agencies detect most foodborne outbreaks through complaints from consumers, very little is known about the availability and use of consumer complaint surveillance systems for foodborne illness.

This study evaluates the use of consumer complaint surveillance systems by local health departments (LHDs) in the United States and their practices and policies for investigating complaints. Data for this study were collected through two Web-based surveys based on a representative sample of LHDs in the United States.

This study found that an estimated 81 percent of LHDs use a complaint-based surveillance system. Of those that did not have a complaint system, 64 percent reported that the state health department or another agency ran their complaint system. LHDs collect a wide variety of information from callers through their complaint systems, including the food intake history. However, most of the LHDs do not store the information in an electronic database. Outbreak rates and complaint rates were found to be positively correlated with a Pearson's correlation coefficient of 0.38. Complaints were the most common outbreak detection mechanism reported by respondents, with a median of 69 percent of outbreaks during the previous year found through complaints.

Complaint systems are commonly used in the United States. Increasing the rate that illnesses are reported by the public and improving investigation practices could help increase the number of outbreaks detected through complaint surveillance.

6.1. Introduction

Foodborne illnesses are an important public health problem, both in terms of the burden of illness and cost to the health care system in the United States.[1, 3, 4] Recently, an report by the Institute of Medicine (IOM) called for improved integration and use of surveillance to help reduce foodborne illness.[89] However, the current surveillance practices of local health departments (LHDs) in the United States are not well-known.

Public health surveillance for foodborne illness is largely carried out at the local level. Efforts to establish national uniformity for pathogen specific surveillance have been made. The Centers for Disease Control and Prevention regularly publishes lists of nationally reportable diseases. Additionally, networks for nationwide reporting of laboratory results have been established.[39] However, many illnesses are not reportable and very little work that looks at other forms of surveillance for foodborne illness has been done.

Consumer complaint-based surveillance has the ability to detect foodborne outbreaks from a variety of pathogens including the majority of non-reportable illnesses.[78] Although these systems have been shown to be effective tools for foodborne illness surveillance, little is known about their use by LHDs. In 2002, a survey of state health officials was carried out in regard to their foodborne disease surveillance and investigation. Authors reported on the ability of state health departments to collect laboratory reports for enteric disease and use of syndromic surveillance, but no information was ascertained on consumer complaint surveillance for foodborne

illness.[58] Without this knowledge, it is difficult to carry out suggestions from the IOM and others to strengthen and better integrate surveillance.

The objective of this project was to survey a representative sample of LHDs to determine the prevalence of consumer complaint surveillance systems in the U.S., assess reasons why this type of surveillance is not used, amount of outbreaks found through complaint surveillance, and practices and policies used for investigation complaints.

6.2. Methods

The National Association of County and City Health Officials (NACCHO) provided a random sample of 500 LHDs, drawn from the all LHDs in the United States. A stratified random sampling design was used. The size of the population served by the LHD was used as basis for stratification. LHDs were split into seven categories based on size of population served: less than 25,000; 25,000 to 49,999; 50,000 to 99,999; 100,000 to 249,999; 250,000 to 499,999; 500,000 to 999,999; and 1,000,000 or greater. The study defined an LHD as an administrative or service unit of local or state government concerned with health and carrying some responsibility for the health of a jurisdiction smaller than the state, the same definition used in NACCHO's profile studies.[90] LHDs in the largest population category were over sampled due to having fewer agencies in this stratum.

Appropriate sampling weights were applied for developing national statistical estimates from the survey. These estimation weights account for the fact that the survey was

administered to a sample of LHDs. They also account for over sampling of LHDs with large population sizes and differences in non-response rates among LHDs in each population size category.

Two self-administered Web-based surveys were conducted for this study, using the web-based survey software Survey Monkey. The first survey (Appendix B) was conducted to obtain general information on LHD surveillance practices and contact information for a food program manager or person in charge of foodborne illness investigations because a preexisting list of such contacts was not readily available.

A link to the first survey was sent via e-mail to the top executive or designated alternate at each LHD in the sample on April 12, 2010. This survey assessed if the LHD had a complaint surveillance system established. The definition of a complaint-based surveillance system was any system that collected information from the public suffering from foodborne illness that they attributed to a particular food establishment, food product, or event. The survey also asked two sets of follow-up questions. The first set of questions was for those who said they have a complaint-based surveillance system. For those LHDs that did not have a complaint-based surveillance system, the questionnaire contained a series of questions about reasons for not having a system. A total of 307 LHDs responded to the first survey, an overall response rate of 61 percent, with over 50 percent response in each population strata.

The second survey (Appendix C) consisted of a more detailed questionnaire, administered to those 190 LHDs that had agreed to participate and provided a contact. This survey included more detailed questions on the LHD practices and policies in regard to complaint surveillance system and foodborne illness investigation. Respondents were also asked about number of outbreaks and complaints from the previous year. The response rate for the second survey was 49 percent.

Questions from the first survey were weighted by the population size category served by the LHD when appropriate. Statistics from the second survey were not weighted since they were based on data from only a sub-sample of the first survey. A complaint rate was calculated by taking the number of complaints in the last year and dividing by the population of that jurisdiction. Outbreak rates were similarly calculated. An outbreak per complaint rate was also calculated by dividing the number of outbreaks in the past year by the number of complaints received by the LHD. Associations between outbreak rates and complaint rates were measured using Pearson's correlation coefficient. The exact phrasing of survey questions presented in tables and results were shortened in some cases from what was in the original survey.

6.3. Results

An estimated 81 percent (95% CI, 76%-86%) of LHDs have a complaint-based surveillance system (Table 1). The proportion with a complaint-based surveillance system varied among LHDs in different categories of size of population served. LHDs in the smallest population categories have the lowest percentage of agencies with a

complaint system at 76 percent, while significantly greater proportion (>92%) of LHDs in the largest population categories have such systems (Table 1). Overall, LHDs reported that their complaint surveillance system was responsible for detecting 69 percent of foodborne outbreaks. LHDs in the largest population category reported that 85 percent of outbreaks in the past year were discovered through a consumer complaint surveillance system.

Almost all LHDs with a complaint system collected contact information for the complainant and the information on the establishment or product he or she suspected caused their illness (Table 2). Ninety-seven percent of agencies collected information on symptoms a caller experienced, and 95 percent collected onset times of illness. A large majority, 85 percent, also collected food histories.

Overall, 16 percent LHDs did not have a consumer complaint system for foodborne illnesses. However, 32 (64%) of these agencies responded that the state or another LHD collects consumer complaints for their jurisdiction. The leading reason for not using a complaint-based surveillance system was the cost of surveillance system; 60 percent of agencies with no existing complaint system would use an electronic complaint database if one was made available at no charge. Lack of resources (28%) and lack of personnel (22%) were the next leading reasons. Only a small proportion of LHDs (4%) do not use a complaint surveillance system because they do not feel it is effective in detecting foodborne outbreaks.

Forty of the LHDs that responded to the survey serve a jurisdiction with a population smaller than 100,000 people. In these 40 LHDs, 34 (85%) reported having no outbreaks in the past year. The percentage of LHDs reporting no outbreaks in the past year decreased by population size served (Figure 1). LHDs in the three largest population categories reported that a median of 1.5, 2, and 7 outbreaks in the past year respectively (Table 3). Of all outbreaks detected by LHDs, 69 percent (95% CI, 57-81) were discovered through complaint surveillance (Figure 2).

There was a wide range in the number of complaints received by each LHD. The median complaint rate for responding LHDs was 21 complaints in the past year per 100,000 people (Table 3). However, rates of complaints were inversely related to population size. LHDs in the smallest population category had the highest rate of complaints (46 per 100,000 population) while the largest two population categories reported the lowest rates of complaints (9 per 100,000 population and 14 per 100,000 population respectively) (Table 3).

There was also a wide range in the number of outbreaks detected per 1,000 complaints received. Outbreaks per 1,000 complaints rose as population size rose, from zero for LHD representing populations less than 100,000 to 36 outbreaks per 1,000 complaints for the largest population category (Table 3). A significant relationship existed between outbreak rates and complaint rates reported by LHDs, with a Pearson's correlation coefficient of 0.38 ($p=0.0004$).

Median outbreaks per 1,000 complaints appeared to be higher for agencies with an electronic database when compared to those without (Figure 3). The difference in medians in the largest 3 population groups were borderline significant, $p=0.10$. Also agencies that had an electronic database were more apt to also have a mechanism to review complaints for common exposures, 83% versus 57% ($p=0.009$) compared to those without.

Ninety percent of LHDs identified foodborne outbreaks through complaints about food establishments; whereas 81% identified outbreaks through a provider report of a reportable illness (Table 4). Laboratory surveillance played a smaller role with 61 percent of jurisdictions reporting detecting outbreaks through this method. Consumers were also able to report complaints through a variety of methods with direct calls to the LHD being the most common with 98 percent of agencies reporting this as a method to receive complaints.

Basic information about outbreaks was collected by most of the LHDs with foodborne illness surveillance systems. Eighty-seven percent collected food intake history; 80 percent of which collected a three-day history, 13 percent collected less than a three day history, and 7 percent collected information for over three days (Table 4). Seventy-nine percent of LHDs sometimes collected stool samples from complainants. Almost every LHD (99%) investigated at least some complaints, with the most common follow-up method being a sanitarian contacting the establishment (Table 4).

6.4. Discussion

The results showed that the majority of LHDs have a complaint-based surveillance system. Additionally, many of those that do not themselves have a complaint system are covered by a state agency or another local health agency. The definition that was used for a complaint surveillance system was broad and reflected the lack of prior knowledge on the uses of such systems by LHDs. An important aspect of this survey was ensuring that all agencies that collect complaint information from the general public to aid in foodborne disease investigation were captured. The results showed that LHDs most collected the core information needed to successfully investigate foodborne outbreaks. The importance of complaint surveillance systems was also illustrated by the large role it played in detection of foodborne outbreaks by LHDs. More LHDs reported this as a detection mechanism for outbreaks than laboratory surveillance, which may also reflect a lack of access to laboratory surveillance among smaller agencies.

The correlation between outbreak rates and complaint rates suggested that those agencies with higher complaint rates also had higher outbreak rates. This could be for a variety of reasons. Information in the second survey indicated that a large proportion of outbreaks were found through complaints and correspondingly, higher complaint rates were associated with higher outbreak rates. Those that collected more complaints also may have a more developed system for investigating and following up on potential cases. The existence of an electronic database and mechanisms to review complaints for common exposures also led to more outbreaks being detected per complaint in larger jurisdictions. This supports the suggestion that proper investigation of complaints is necessary to detect outbreaks.

The complexity of foodborne illness surveillance and investigation makes characterizing what each LHD does through a survey difficult. A few areas were targeted for questioning that are likely to help improve investigations: if follow-up was carried out, if necessary information was collected, availability of stool samples for testing, and if information was shared.[58, 91] Most LHDs surveyed indicated that complaints are routinely investigated either through follow-up with the establishment by a sanitarian or follow-up with the complainant by LHD staff. Most also reported the ability to collect stool samples from complainants. Consistent application of these practices should lead to an enhanced ability to detect a foodborne outbreak.

This project had several limitations. The second survey had a low response rate. This survey provided useful information but may not be representative of the practices of LHDs. The second limitation was that all information was self-reported. Although the questions were regarding practices and policies of the LHD, there could have been issues with recall and need to interpret practices that did not exactly fit the questions. Thus, results may portray a more optimistic picture of complaint-based surveillance practices and investigations than would be determined by a more rigorous evaluation.

Despite these limitations, this survey shows that the much of the surveillance infrastructure needed for foodborne illness outbreak detection exists. To maximize the utility of this existing infrastructure, it is important to use it in the manner most likely to enable detection of outbreaks. It appears that many jurisdictions collect detailed information from each caller. However, a large proportion of LHDs did not find any

outbreaks in the previous year. For health agencies covering an area with a population greater than 1,000,000 people, there were still 20 percent of agencies reporting no outbreaks. There were likely outbreaks that occurred in these jurisdictions that were never detected, indicating the need for more work to refine the use of foodborne illness surveillance including the use of complaint surveillance. The success of complaint systems not only depends on collecting the proper information but being able to review and act on incoming data. Agencies with an electronic database reported a better ability to review incoming data for similar exposures. This is necessary to identify possible outbreaks in data. Increased use of electronic databases, aggregating data across small jurisdictions, and applying uniform practices of investigation recommended by the Council to Improve Foodborne Outbreak Response could greatly increase the utility of foodborne illness complaint systems.[44]

Acknowledgments

Funding for this study was provided through a grant from the National Center for Food Protection and Defense, grant number 2007-ST-061-000003-02.

6.5. Tables

Table 1: Percentage of health departments with a consumer complaint surveillance system by population of health department jurisdiction

Population	Percentage with system (95% CI)
<25,000	76 (67, 86)
25,000-49,999	83 (72, 93)
50,000-99,999	84 (73, 95)
100,000-249,999	82 (71, 94)
250,000-499,999	86 (75, 98)
500,000-999,999	96 (89, 100)
≥1,000,000	92 (81, 100)
Total	81 ^a (76, 86)

^a Weighted percentage

Table 2: Percentage of health departments that collect key information from complainants from those that said they had a complaint system

Information	Percent that Collects (95% CI) ^a
Complainant Contact Info	99 (99, 100)
Suspected establishment/product	99 (97, 100)
Symptoms	97 (95, 100)
Onset of time illness	95 (92, 98)
Sought health care	93 (89, 97)
Info on other ill individuals	92 (88, 98)
Number ill in group	89 (84, 94)
Food history	85 (80, 91)
If complaint had a stool sample tested	82 (76, 88)
Number exposed in group	81 (75, 87)

^a n=257, 20 respondents who answered that they did have a complaint system did not answer this series of questions

Table 3: Numbers of outbreaks, outbreaks detected by complaints, and rates of outbreaks and complaints in the past year by health agency by population category

Population	Number of Agencies	Median Outbreaks (1Q, 3Q)	Median Outbreaks per 1,000 Complaints (1Q, 3Q)	Median Rate of Complaint per 100,000 Population (1Q, 3Q)
<25,000	13	0 (0,0)	0 (0,0)	46 (20, 100)
25,000-49,999	17	0 (0,0)	0 (0,0)	26 (15, 41)
50,000-99,999	10	0 (0,0)	0 (0,20)	32 (9, 46)
100,000-249,999	12	0.5 (0, 2.5)	4 (0, 55)	26 (15, 114)
250,000-499,999	12	1.5 (0.5, 3.5)	18 (3, 52)	25 (15, 63)
500,000-999,999	9	2 (1, 5)	20 (3, 167)	9 (6, 18)
≥1,000,000	16	7 (4, 16.5)	36 (19, 48)	14 (9, 18)
Total	89	0 (0, 3)	7 (0, 43)	21 (11, 47)

Table 4: Summary of local health department use of consumer complaint surveillance systems

Question	n	% yes
How are outbreaks detected in your jurisdiction?	89	
Complaints about a food establishment		90
Complaints about an event		88
Provider report of reportable illness		81
Report of illness from an institution		64
Isolate submission from lab		61
Clinician report of non-reportable illness		40
Report of complaints from food establishment		40
Other		9
How are consumer complaint received?	87	
Staff member takes complaint over the phone		98
E-mail		75
In person		72
Voicemail		69
Web-based reporting form		40
Other		5
Is a standard set of questions asked of each complainant?	86	88
Is a food history taken?	85	87
How long is the food history?	74	
3 days		80
Less than 3 days		13
More than 3 days		7
Is information collected stored in an electronic database?	86	43
Are complaints investigated?	84	99
Who investigates complaints?	83	
Environmental Health, sanitarian inspects establishment		88
Health official/epidemiologist contacts caller		84
Other		17
Is there a system to review individual incoming complaints for similar exposures?	86	69
Are stool samples collected from callers?	86	
Sometimes		79
Never		20
Always		1
Is information collected shared with the state health department?	86	69
Is information shared with other local agencies?	86	55

6.6. Figures

Figure 1: Percentage of agencies that reported not having a foodborne outbreak in the past year by population size

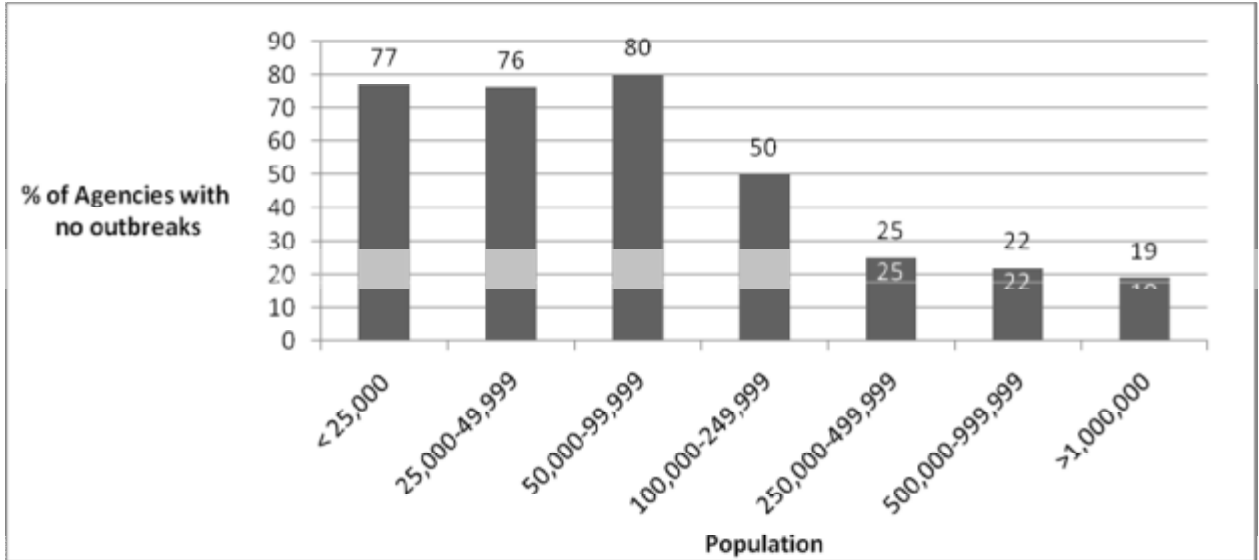


Figure 2: Percentage of outbreaks found through complaint based surveillance systems with 95 percent confidence intervals

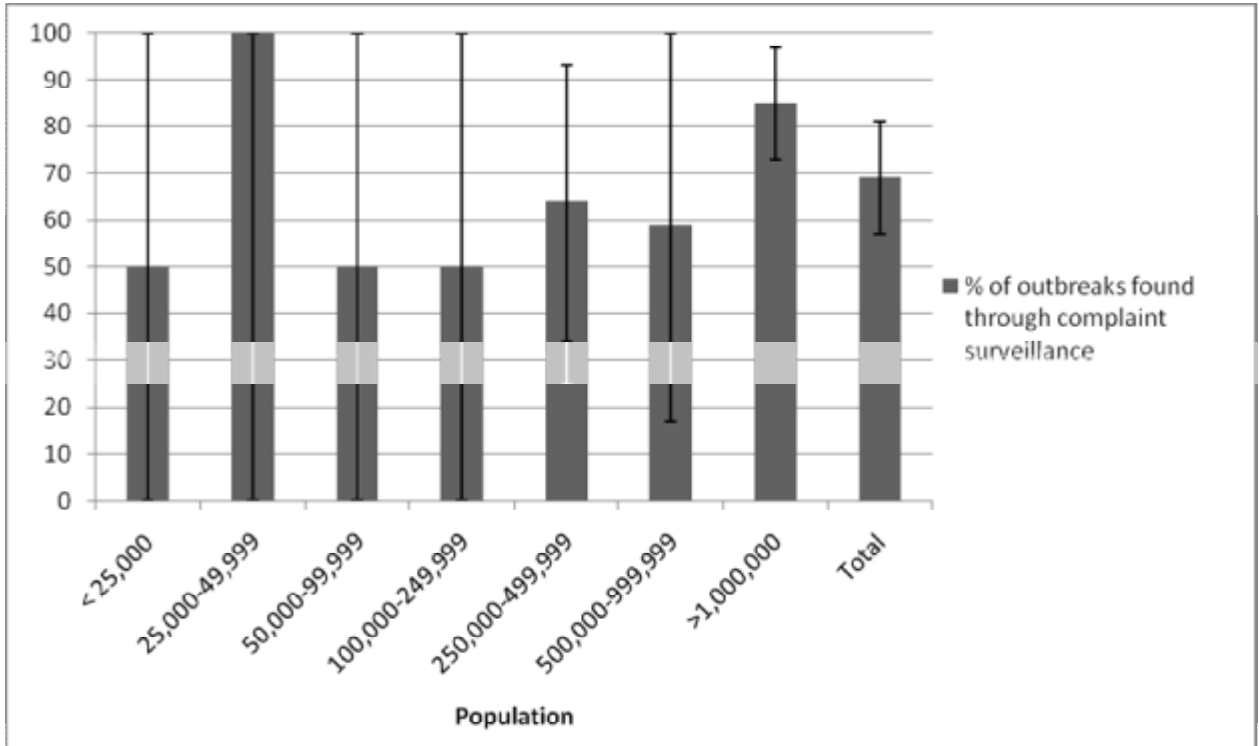
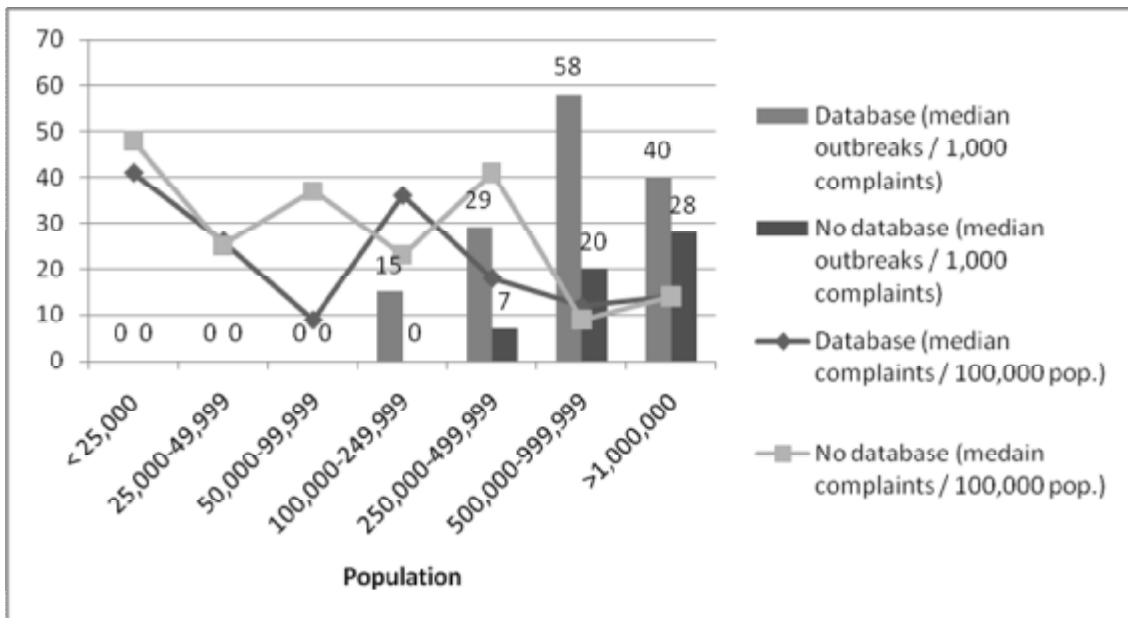


Figure 3: Comparison of median outbreak rates and outbreaks per complaint rates by population category and if the agency reported having an electronic database or not



Chapter 7. Conclusion and Implication of Future Research

The goal of this dissertation was to propose methods to improve foodborne illness surveillance in the United States. In particular, focusing on complaint based surveillance which has received very little attention, but has been suggested as a powerful surveillance tool for foodborne illness. This dissertation sought to enhance foodborne illness surveillance through two main aims. The first was the systematic characterization and evaluation of a robust statewide complaint based surveillance system. This included examining characteristics that distinguished outbreak related complaints to non-outbreak related complaints and description of the types of outbreaks detected by complaint systems. Additionally, a survey of local health departments was carried out to determine the current use of consumer complaint base surveillance systems in the United States. The second aim was the development of new methods to maximize the utility of using consumer complaint data for outbreak detection. There has been little research done looking at the use of complaint based surveillance systems for foodborne outbreak detection especially focusing on methods to better detect clusters of illness in incoming data.

In manuscript one the complaint surveillance system in Minnesota was evaluated and described. The majority of confirmed foodborne outbreaks during this period in Minnesota were found through complaint surveillance. This included almost all non-reportable illnesses and several outbreaks due to reportable illnesses traditionally found through pathogen specific surveillance. However, success in finding illnesses is not the only measure in determining the appropriateness of a surveillance system. Evaluation of

a surveillance system is a complicated task that must balance several factors. The Centers for Disease Control and Prevention has provided some guidance on criterion that should be considered when evaluating a surveillance system, including timeliness of the system, sensitivity, specificity, acceptability, cost, representativeness, flexibility, and usefulness [46, 92]. The analysis presented in this dissertation could not address all of these issues. Even though all criteria to evaluate a surveillance system could not be directly addressed, it is still important to discuss the multiple factors that make complaint based surveillance a beneficial tool for foodborne disease outbreak detection. It is also important to judge each criterion in relation to the purpose of the surveillance system, the most important criteria to consider for a foodborne illness surveillance system is likely much different than for surveillance of other diseases.

Timeliness is one of the most important factors in foodborne disease surveillance. Delays of notification of illness have been cited as reasons why potential outbreaks are not investigated. Additionally, successful secondary prevention of illness depends on identifying the cause of illness and removing it from the marketplace when possible. In this dissertation the timeliness of a complaint based surveillance system was compared to that of pathogen specific surveillance. It was found that the time between onset of illness and report to the health department was significantly shorter for cases reported through complaint based surveillance. Consumer complaint surveillance allows the public to directly contact the health department allowing much faster report of illness and initiation of investigations. Shorter time lags between illness onset and start of investigations

would also aid in determining common exposures by improving the likelihood that the ill individual remembers items they ate prior to illness.

Sensitivity and positive predictive value are also important metrics for a surveillance system. These metrics also tend to be low for complaint based surveillance systems. In fact information about these metrics was not presented in the evaluation of Minnesota's complaint system. They were only presented in context of examining screening tools proposed in manuscript two and three. Sensitivity and ppv are difficult to measure for a foodborne illness surveillance system. The reason for this is that only a small fraction of complaint calls turn out to be caused by confirmed foodborne outbreaks. A sensitivity cannot be calculated properly without some type of screening mechanism to determine which calls need to be investigated. However, it is important to consider the amount of calls that are related to outbreaks in a complaint system. In Minnesota, only seven percent of complaint calls over a 7 year period were related to a confirmed outbreak. As noted complaint systems are apt to contain quite a bit of additional information. The key is screening out calls using questioning of common exposures, or other proposed methods to improve sensitivity and ppv.

Information presented in this dissertation can lead to improvements in sensitivity and positive predictive value in the use of complaint based surveillance. The main goal of health practitioners is to better target individuals for follow up. In this respect the work in this dissertation points out several markers that are currently used to target investigations and proposed several new methods to focus investigations. Current

practice which should be encouraged in jurisdictions not currently doing it is to investigate all complaints where there is a clear common exposure. The number ill in a call was found to be associated to be an outbreak related complaint. Additionally, those that were related to confirmed outbreaks seemed to be much better at identifying their exposure. This indicates given additional information, such as numerous ill individuals or multiple households of ill individuals, people are very adept at determining their cause of illness. It is essential that complaint systems be able to collect this information from callers to immediately follow up on. Proposed methods to screen for pathogens of possible clusters of illness could also be used to better utilize complaint surveillance data. Manuscripts two and three detail new methods that could be used to identify callers of interest in complaints data. One method focuses on etiology of a caller's illness, the other focuses on unusually high numbers of calls. Using both methods to screen incoming calls would help to improve the positive predictive value for complaint surveillance systems.

Other issues for surveillance systems are costs and acceptability. In this dissertation direct costs of a complaint surveillance system were not investigated. It is likely quite costly to employ multiple individuals to regularly collect information for complaint callers and follow up on possible outbreaks. However, this is not an additional cost to health departments. A main function of public health departments is foodborne illness surveillance and prevention. As evidenced by the national survey presented most health departments in the United States have some type of complaint surveillance system. Modifying existing system to collect the proper data to investigate outbreaks would not

be costly. This also addresses issues of acceptability. Very few health departments surveyed reported that complaint surveillance systems were not used because they were ineffective in finding outbreaks. Complaint based surveillance systems are well accepted they are just not used as effectively as they could be.

The last issue and perhaps the most important when considering the value of a surveillance system is usefulness. To be useful a surveillance system needs to accomplish the task it was designed for without duplicating the efforts on another surveillance system. As mentioned foodborne illness surveillance is also done using pathogen specific surveillance. These systems are not able to detect outbreaks that are due to non-reportable pathogens. They also are limited by underreporting. Complaint based surveillance systems provide an excellent complement to pathogen specific surveillance allowing detection of non-reportable pathogens as well as providing extra case finding for some reportable pathogens. Complaint based surveillance also is not as affected by underreporting since it allows direct report of illnesses from the general public. Manuscript one and results from the survey of local health departments clearly illustrate the ability of complaint based surveillance systems to find outbreaks. This type of system is both useful in performing its task and unique in its ability to find outbreak due to a large variety of pathogens. It was also evident that the utility of these systems could be improved with new methods to examine incoming data.

The second aim of this dissertation was to develop new methods to aid in the effective use of complaint surveillance systems and increase the ability to detect outbreaks. As

mentioned this is closely related to the effectiveness and utility of a surveillance system. Research in automated cluster detection and profiling of outbreaks has been done in other areas. Cluster detection methods have been applied to pathogen specific surveillance, namely detection of *Salmonella* clusters that shared similar serotypes. Additionally, various methods have been applied to syndromic surveillance to detect increases in possible cases. These methods had never been extended to complaint data. It was thought that applying developing such methods for complaint surveillance would lead to increased utility of such surveillance systems. In addition to improving public health practice developing these methods also leads to new research directions on ways to improve surveillance.

Findings from this dissertation have the potential to have significant public health impact and policy implications. This dissertation was novel in examining a potentially underutilized source of data for detection of foodborne outbreaks. Complaint surveillance provides a powerful tool to health departments to complement pathogen specific surveillance in detection of foodborne outbreaks. Improvement of foodborne surveillance in the United States could lead to large reductions in foodborne illnesses in the United States. There is currently interest in improving food safety in the United States in part by focusing on outbreak detection methods and better integration of various data sources [89, 93]. Extending the work from this dissertation further could lead to changes in surveillance practices and improvements in foodborne disease investigations.

This dissertation provides a framework to develop an improved foodborne illness surveillance system in the United States. The worth of complaint based surveillance in foodborne disease investigations has been shown. This dissertation also highlighted information that must be collected by complaint systems and best practices in investigation of outbreaks. In addition to this new methods to better utilize incoming data have been developed. These methods will become ever more important if the vision of a large connected surveillance system comes to fruition. Increased data volume from larger surveillance system will require methods to help screen data. Improvement of complaint systems could lead to vast benefits in the amount and speed that outbreaks are detected. Existing infrastructure could also be leveraged to enhance surveillance and foodborne outbreak investigation practices on the local level as well as allowing scalability to aggregate and analyze data on a regional and national level. A combination of more research in utilizing surveillance data and putting into practice improved methods will help address the larger goal of reducing the burden of foodborne illnesses. A combination of pathogen specific surveillance and complaint based surveillance will need to be used to address the issue of foodborne illnesses.

7.1. Key Findings

Throughout this dissertation various findings have been presented. Below a number of key findings are highlighted with discussion on how they are related to the larger goal of enhancing foodborne surveillance in the United States.

Complaint based surveillance is able to detect foodborne outbreaks caused by a variety of pathogens, including some reportable illnesses. This type of surveillance has been cited previously as the only method to detect non-reportable illnesses and new and emerging illnesses. The analysis presented supported the fact that almost all foodborne outbreaks that were due to non-reportable illnesses in Minnesota were found through their complaint surveillance system. Given that only a small percentage of the general public will seek medical attention for their foodborne illness there is really no other method that these illnesses would have been picked up. Additionally, manuscript one presented data that showed that a 25% of *Salmonella* outbreaks were also found through complaint surveillance. This is an important finding that refutes much of the current thinking that consumer complaint surveillance would only find norovirus outbreaks. Pathogen specific surveillance is often seen as the most important source of data for detection of outbreaks, especially for those pathogens of most interest. Finding that complaint based surveillance can also detect reportable pathogens illustrates the importance of using multiple methods of surveillance for detecting outbreaks.

Times between onset of illness and report were much shorter for individuals reporting through complaint based surveillance than those through pathogen specific surveillance. To be picked up by pathogen specific surveillance an individual needs to visit a physician and have a stool sample collected for testing. This requirement is hurt by low numbers of individuals that visit a physician for a foodborne illness and also can result in long reporting times. After a stool sample is collected it needs to be sent to a lab and tested, after a result is known then it is reported to the health department. These additional steps

between becoming ill and being reported to the health department explain the discrepancy in times. In complaint based surveillance individuals report directly to the health department. This allows much faster reporting times but it also results in an inability to make an exact diagnosis. Those calling suspect that they have a foodborne illness but this cannot be confirmed unless an investigation is carried out. Use of consumer complaint based surveillance can lead to faster initiation of investigations.

Outbreak related callers appeared to be able to identify the proper source of their illness more readily than non-outbreak related callers. This was apparent in the examination of incubation periods of reported illnesses between outbreak and non-outbreak callers. Non-outbreak related callers had much shorter incubation periods suggesting that they often suspected the last place they ate as causing their illness. This corresponds with current literature in the area that individuals have difficulty in determining the source of their illness if a foodborne source is suspected. This inability to determine the correct source of exposure indicates the necessity for collecting a food history from most individuals. However, it is also interesting to point out that the incubation periods for outbreak related calls matches better with the incubation periods for the most common causes of foodborne illnesses. This indicates that callers that were related to confirmed outbreaks often correctly identified the exposure of their illness. Although the general public is not well versed in determining causes of their illness, it appears when given other information such as other sick companions they are very good at identifying a common exposure.

The most important predictor if a call was associated to an outbreak or not was the number of ill individuals associated with the call. This was likely due to investigation practices of the health department but it still provides insight on calls that should always be investigated. A larger number of ill individuals, especially when from separate households are more apt to correctly identify the source of their illness. MDH always initiates investigations when a common source of exposure is clearly indicated which in most instances corresponds to a higher number of ill. Increased investigations will lead to more outbreaks being found. After examining investigation practices it was clear that it was not necessarily the number ill driving the relationship to outbreak associated calls, but the ability of the health department to identify a common exposure when more people were ill. This leads to the conclusion that health departments must collect enough information to ascertain common exposures among ill individuals to follow up on those calls that are almost assuredly due to a foodborne outbreak.

A predictive model for *Salmonella* was developed and shown to be able to screen callers for *Salmonella*. More work in this area needs to be done to develop better predictive models and better understand sporadic illnesses. However, it is a tool that could help enhance the utility of complaint based surveillance. A large issue with using complaint surveillance is the large volume of incoming calls. Investigating all calls is not feasible for most health departments. Increased use of models to screen for pathogens could help identify callers that require additional follow up possibly leading to detection of more outbreaks.

Cumulative sum methods can be used to identify clusters of illness using complaint surveillance data. They appear to be best able to identify outbreaks with multiple individual complaints and norovirus outbreaks. Their use could have an impact on disease surveillance in several respects. They provide another method to identify calls for further investigation. They also allow health departments that do not collect enough information or do not regularly review incoming calls to have a way to flag potential clusters of illness. This may become more important as more complaints are received through automatic mechanisms such as the internet versus from direct reporting to a health department official. Additionally, these algorithms may allow complaint surveillance to act as an early warning system for rises in foodborne illnesses. A general rise in call volume could indicate an increase in illness in the community, triggering warning to establishments most vulnerable to foodborne outbreaks.

The majority of local health departments have some type of surveillance system to collect consumer complaints of foodborne illness. They also reported to ability to collect stool specimens and collection of a variety of information needed to investigate potential clusters of illness. However, the number of outbreaks per complaints in a jurisdiction varied widely, suggesting some agencies have a better ability to investigate and detect outbreaks. The majority of outbreaks found by local health departments is found through complaint based surveillance, suggesting that strengthening this system would lead to detection of more outbreaks.

7.2. Strengths and Limitations

There were strengths and limitations to the data source used in this study. The use of Minnesota Department of Health's foodborne illness database, allowed the pursuit of several research questions that had not been previously addressed. Very few studies have been carried out specifically examining consumer complaint surveillance for foodborne outbreak detection. One previous study found that complaint systems may have the ability to detect outbreaks but was not able to link outbreaks that occurred in the jurisdiction to complaint entries in the database [45]. Using MDH's complaint database provided the ability to link complaints to confirmed outbreaks. MDH also actively investigates a wide range of foodborne illnesses allowing the detection of more outbreaks than most state health departments [8]. The ability to link complaints to confirmed outbreaks allowed for the characterization of complaint related calls and analysis of factors most important to collect to identify outbreaks. It also allowed the creation of predictive models and analysis of types of outbreaks found through complaints. However, using this surveillance data had its limitations. Not all outbreaks that were found through complaint surveillance were linked in the database. Also there are likely quite a few outbreaks among callers that were never investigated. This could have affected the results due to misclassification of some outbreak related calls as non-outbreak related. Even with this possible bias the results found are still useful for health departments. Using differences found in complaint calls that led to confirmed outbreaks will still help health departments focus investigations and collect better information to allow for identifying common exposures among callers.

Data quality may have varied over the course of the study period. Approximately 25% of the calls each year are received by a local health department then passed on the MDH, the remaining are received directly by MDH. The majority of calls received by MDH over the study period were answered by one person using the same questionnaire, so were likely to have been very consistent. Callers to local health departments could have been interviewed by a variety of personnel. Local health department did use the same questionnaire, but there could have been more errors or omissions in collected information in the early periods of the complaint surveillance system as people were getting accustomed to the new system. Additionally, lab methods improved over the study period allowing better detection of some pathogens such as norovirus, resulting in more outbreaks of unknown etiology in earlier years of the study. Large differences in missing information were observed for different years under study. However, there was more difficulty in linking outbreaks in earlier years. A few outbreaks that were noted as being found through complaints could not be found in the complaint database for the earlier years of the study. Being that the number of these missing outbreaks was small, only it is not likely that this would have had a great impact on the results.

Another limitation of using consumer reported complaint data is the potential reporting bias that may occur in certain situations. Complaint based surveillance systems take complaints from the general public, reporting illnesses that they attribute to a certain product or food establishment. It is plausible that after outbreaks are reported in the media there could be increased complaints due to individuals wanting to report that they ate the implicated product. If these large outbreaks tended to be associated with certain

pathogens this could skew the findings of the types of outbreaks detected through complaint surveillance towards those pathogens. This did not appear to be a large problem in the 7 years of data that was analyzed. Very few outbreaks were large enough to warrant media attention. Additionally, the timing of the media reports is usually far after the health department investigates and determines a cause of illness. This reporting bias may also have an impact on cusum methods that look for spikes in calls. Increase call volume due to media reports could trigger flagged weeks. This may be a problem if a large number of these calls are not related to the outbreak. If they are related to the outbreak finding more cases can actually help shed more light on the cause of illness and possibly other contaminated products. Again due to the low amount of outbreaks that reach the level of media attention this is not likely a large issue. In fact increased publicity on the use of complaint systems and the need for the public to report illnesses would help the ability of a system to detect more outbreaks.

Survey work done in this dissertation was the first of its kind to look at complaint surveillance use at the local health department level. Very little is known about foodborne illness surveillance practices by local health departments. This is a large gap in knowledge since most foodborne illness investigation is run on the county level. However, the fractured state of foodborne illness surveillance also makes surveying health departments difficult. A broad definition of a complaint system had to be used to capture as much information as possible from respondents. What health departments consider to be a complaint system could be anything from a health department that occasionally gets a report about a restaurant and does little to act on it to a health

department that has a system to collect calls store them electronically, and follows up with individuals to determine exposures. The lack of a basic minimum to what constitutes a complaint surveillance system makes work in this area difficult. The work done in this dissertation attempts to address what needs to be done to constitute a working complaint surveillance system.

7.3. Implications for Future Research and Foodborne Surveillance Practices

Findings from this dissertation lead to several possible avenues for further research that would allow the better understanding of foodborne illnesses in general and improvement on the detection methods that have been proposed. As mentioned the evaluation and development of methods to use complaint surveillance data also lead several practical improvements that can be made to foodborne illness surveillance. In this section extensions of the research started in this dissertation and implementation of new complaint surveillance practices are discussed.

One vein of research that could be pursued would be to gain a better understanding of sporadic illnesses among complaint callers. Determination of the pathogen that caused a caller's illness was largely dependent on them being associated to an outbreak. Lab tests were seldom carried out on individuals that were not suspected of being part of an outbreak. Although, illnesses among those that were not identified as being part of a confirmed outbreak likely are caused by similar pathogens it has never been shown to be true. Knowing causes of sporadic illnesses would allow a better understanding of the

types of illnesses complaint systems can detect and also allow us to estimate to amount of foodborne illness that can be attributed to different pathogens.

Prediction models for other pathogens also need to be developed. This will be possible as an extension of research looking into sporadic causes of illness. In addition to determining a the etiology of a caller's illness this information can also be used in the development of more predictive models and the improvement of the *Salmonella* prediction model presented in this dissertation. This could include looking for other predictors of illness beyond symptoms.

Information discovered in this dissertation could be used to develop a platform for increased complaint surveillance. Such systems could lead to vast improvements in outbreak detection. Electronic databases provide a better means to store and retrieve data collected by surveillance systems. They also allow the application of automatic detection algorithms like the cusum method developed in this dissertation. A natural extension of the work presented here is the development of a suite of tools that can be accessed by local health departments. This would include an electronic database and standardized questionnaire for complaint surveillance. Included in the database could be tools to help identify which calls should be investigated. Standardizing collection and data formats at the local level would also allow information to be shared more easily across various jurisdictions. Work in this area would be aided by the fact that most public health jurisdictions have some form of complaint based surveillance system. Along these lines

work on expanding complaint based surveillance systems could lead to the integration of various disparate systems for improving food safety.

Currently, there is no mechanism to incorporate information from distribution networks for food products, data food establishments, and recalls of contaminated products. A natural extension to complaint based surveillance would be to link these sources of information to be able to link complaints to recalls or contaminated products and tracing them through various distribution nodes. Being able to uniquely identify all establishments through tax identification numbers or other registration information could lead to better automated detection of common exposures. This would necessitate that health departments take detailed food histories and record them in a way that could be easily used such as an electronic database. Developing test models of these systems could illustrate their power to detect outbreaks. Additionally, if distribution of products could be linked to establishments than outbreaks of products that are widely distributed could also be more easily identified.

Increasing research of foodborne illness and applying newly developed methods and practices described in this dissertation could lead to improved secondary prevention effort to reduce the burden of foodborne illnesses. Most of the infrastructure to develop a robust complaint surveillance system already exists at state and local public health agencies in the United States. Improvements to how data is used and shared would not be an overly burdensome task once the initiative to change is undertaken. Faster detection of more outbreaks would not only lead to the early removal of contaminated

product but greatly increase the knowledge of the spread of foodborne illness and methods to prevent contamination.

Bibliography

1. Motarjemi, Y. and F.K. Kaferstein, *Global estimation of foodborne diseases*. World Health Stat Q, 1997. **50**(1-2): p. 5-11.
2. Buzby, J.C. and T. Roberts, *Economic costs and trade impacts of microbial foodborne illness*. World Health Stat Q, 1997. **50**(1-2): p. 57-66.
3. Buzby, J.C. and T. Roberts, *The economics of enteric infections: human foodborne disease costs*. Gastroenterology, 2009. **136**(6): p. 1851-62.
4. Mead, P.S., et al., *Food-related illness and death in the United States*. Emerg Infect Dis, 1999. **5**(5): p. 607-25.
5. de Wit, M.A., et al., *Sensor, a population-based cohort study on gastroenteritis in the Netherlands: incidence and etiology*. Am J Epidemiol, 2001. **154**(7): p. 666-74.
6. Kaferstein, F.K., Y. Motarjemi, and D.W. Bettcher, *Foodborne disease control: a transnational challenge*. Emerg Infect Dis, 1997. **3**(4): p. 503-10.
7. Woteki, C.E. and B.D. Kineman, *Challenges and approaches to reducing foodborne illness*. Annual Review of Nutrition, 2003. **23**: p. 315-344.
8. Centers for Disease Prevention and Control, *Surveillance for Foodborne-Disease Outbreaks - United States, 1998-2002*. MMWR, 2006. **55**(SS-10).
9. Pires, S.M., et al., *Attributing the Human Disease Burden of Foodborne Infections to Specific Sources*. Foodborne Pathogens and Disease, 2009. **6**(4): p. 417-424.
10. Altekruse, S.F. and D.L. Swerdlow, *The changing epidemiology of foodborne diseases*. American Journal of the Medical Sciences, 1996. **311**(1): p. 23-29.
11. Kendall, P.A., V.V. Hillers, and L.C. Medeiros, *Food safety guidance for older adults*. Clinical Infectious Diseases, 2006. **42**(9): p. 1298-1304.
12. Koehler, K.M., et al., *Population-based incidence of infection with selected bacterial enteric pathogens in children younger than five years of age, 1996-1998*. Pediatric Infectious Disease Journal, 2006. **25**(2): p. 129-134.
13. Jones, T.F., et al., *A case-control study of the epidemiology of sporadic Salmonella infection in infants*. Pediatrics, 2006. **118**(6): p. 2380-2387.
14. Ethelberg, S., et al., *Risk factors for diarrhea among children in an industrialized country*. Epidemiology, 2006. **17**(1): p. 24-30.
15. Rooney, R.M., et al., *A review of outbreaks of foodborne disease associated with passenger ships: Evidence for risk management*. Public Health Reports, 2004. **119**(4): p. 427-434.
16. Widdowson, M.A., et al., *Outbreaks of acute gastroenteritis on cruise ships and on land: identification of a predominant circulating strain of norovirus--United States, 2002*. J Infect Dis, 2004. **190**(1): p. 27-36.
17. Widdowson, M.A., et al., *Norovirus and foodborne disease, United States, 1991-2000*. Emerging Infectious Diseases, 2005. **11**(1): p. 95-102.
18. Dolin, R., *Perspective: Noroviruses - Challenges to control*. New England Journal of Medicine, 2007. **357**(11): p. 1072-1073.
19. Lopman, B.A., et al., *Viral gastroenteritis outbreaks in Europe, 1995-2000*. Emerg Infect Dis, 2003. **9**(1): p. 90-6.
20. Mounts, A.W., et al., *Cold weather seasonality of gastroenteritis associated with Norwalk-like viruses*. J Infect Dis, 2000. **181 Suppl 2**: p. S284-7.

21. Glass, R.I., et al., *The epidemiology of enteric caliciviruses from humans: a reassessment using new diagnostics*. J Infect Dis, 2000. **181 Suppl 2**: p. S254-61.
22. Dolin, R., *Noroviruses--challenges to control*. N Engl J Med, 2007. **357**(11): p. 1072-3.
23. Rockx, B., et al., *Natural history of human calicivirus infection: a prospective cohort study*. Clin Infect Dis, 2002. **35**(3): p. 246-53.
24. Fretz, R., et al., *Frequency of Norovirus in stool samples from patients with gastrointestinal symptoms in Switzerland*. Eur J Clin Microbiol Infect Dis, 2005. **24**(3): p. 214-6.
25. Billgren, M., et al., *Epidemiology of Norwalk-like human caliciviruses in hospital outbreaks of acute gastroenteritis in the Stockholm area in 1996*. J Infect, 2002. **44**(1): p. 26-32.
26. Lopman, B.A., et al., *Institutional risk factors for outbreaks of nosocomial gastroenteritis: survival analysis of a cohort of hospital units in South-west England, 2002-2003*. J Hosp Infect, 2005. **60**(2): p. 135-43.
27. Tsang, O.T., et al., *Clinical characteristics of nosocomial norovirus outbreaks in Hong Kong*. J Hosp Infect, 2008. **69**(2): p. 135-40.
28. Hook, E.W., *Salmonellosis: certain factors influencing the interaction of Salmonella and the human host*. Bull N Y Acad Med, 1961. **37**: p. 499-512.
29. Olsen, S.J., et al., *The changing epidemiology of salmonella: trends in serotypes isolated from humans in the United States, 1987-1997*. J Infect Dis, 2001. **183**(5): p. 753-61.
30. Marcus, R., et al., *Re-assessment of risk factors for sporadic Salmonella serotype enteritidis infections: a case-control study in five FoodNet sites, 2002-2003*. Epidemiology and Infection, 2007. **135**(1): p. 84-92.
31. Hennessy, T.W., et al., *Egg consumption is the principal risk factor for sporadic Salmonella serotype Heidelberg infections: A case-control study in FoodNet sites*. Clinical Infectious Diseases, 2004. **38**: p. S237-S243.
32. Bryan, F.L., *Hazard analysis critical control point approach to food safety - Past, present, and future*. Journal of Environmental Health, 1999. **61**(8): p. 9-14.
33. Motarjemi, Y. and F. Kaferstein, *Food safety, Hazard Analysis and Critical Control Point and the increase in foodborne diseases: a paradox?* Food Control, 1999. **10**(4-5): p. 325-333.
34. Shiferaw, B., et al., *Prevalence of high-risk food consumption and food-handling practices among adults: A multistate survey, 1996 to 1997*. Journal of Food Protection, 2000. **63**(11): p. 1538-1543.
35. Tauxe, R.V., *Surveillance and investigation of foodborne diseases; roles for public health in meeting objectives for food safety*. Food Control, 2002. **13**(6-7): p. 363-369.
36. Guzewich, J.J., F.L. Bryan, and E.C.D. Todd, *Surveillance of foodborne disease .I. Purposes and types of surveillance systems and networks*. Journal of Food Protection, 1997. **60**(5): p. 555-566.
37. Swaminathan, B., et al., *Building PulseNet International: An interconnected system of laboratory networks to facilitate timely public health recognition and response to foodborne disease outbreaks and emerging foodborne diseases*. Foodborne Pathogens and Disease, 2006. **3**(1): p. 36-50.

38. Swaminathan, B., et al., *PulseNet: The molecular subtyping network for foodborne bacterial disease surveillance, United States*. Emerging Infectious Diseases, 2001. **7**(3): p. 382-389.
39. Gerner-Smidt, P., et al., *PulseNet USA: A five-year update*. Foodborne Pathogens and Disease, 2006. **3**(1): p. 9-19.
40. McNabb, S.J., et al., *Summary of notifiable diseases--United States, 2006*. MMWR Morb Mortal Wkly Rep, 2008. **55**(53): p. 1-92.
41. Green, L.R., et al., *Beliefs about meals eaten outside the home as sources of gastrointestinal illness*. J Food Prot, 2005. **68**(10): p. 2184-9.
42. Swaminathan, B., T.J. Barrett, and P. Fields, *Surveillance for human Salmonella infections in the United States*. J AOAC Int, 2006. **89**(2): p. 553-9.
43. Chalker, R.B. and M.J. Blaser, *A Review of Human Salmonellosis .3. Magnitude of Salmonella Infection in the United-States*. Reviews of Infectious Diseases, 1988. **10**(1): p. 111-124.
44. Council to Improve Foodborne Outbreak Response (CIFOR), *Guidelines for Foodborne Disease Outbreak Resposnse*. 2009, Atlanta: Council of State and Territorial Epidemiologists.
45. Samuel, M.C., et al., *Complaints of foodborne illness in San Francisco, California, 1998*. J Food Prot, 2001. **64**(8): p. 1261-4.
46. Teutsch, S.M. and R.E. Churchill, *Principles and practice of public health surveillance*. 2nd ed. 2000, Oxford ; New York: Oxford University Press. xiii, 406 p.
47. Gargouri, N., et al., *Estimated burden of human Salmonella, Shigella, and Brucella infections in Jordan, 2003-2004*. Foodborne Pathog Dis, 2009. **6**(4): p. 481-6.
48. Martinez, A., et al., *Epidemiology of foodborne norovirus outbreaks in Catalonia, Spain*. BMC Infect Dis, 2008. **8**: p. 47.
49. Hall, J.A., et al., *Epidemiologic profiling: evaluating foodborne outbreaks for which no pathogen was isolated by routine laboratory testing: United States, 1982-9*. Epidemiol Infect, 2001. **127**(3): p. 381-7.
50. Hedberg, C.W., et al., *The use of clinical profiles in the investigation of foodborne outbreaks in restaurants: United States, 1982-1997*. Epidemiol Infect, 2008. **136**(1): p. 65-72.
51. Hope, K., et al., *Syndromic surveillance: is it a useful tool for local outbreak detection?* Journal of Epidemiology and Community Health, 2006. **60**(5): p. 374-375.
52. Kleinman, K., R. Lazarus, and R. Platt, *A generalized linear mixed models approach for detecting incident clusters of disease in small areas, with an application to biological terrorism*. Am J Epidemiol, 2004. **159**(3): p. 217-24.
53. Marx, M.A., et al., *Diarrheal illness detected through syndromic surveillance after a massive power outage: New York City, August 2003*. American Journal of Public Health, 2006. **96**(3): p. 547-553.
54. Sivapalasingam, S., et al., *A multistate outbreak of Salmonella enterica Serotype Newport infection linked to mango consumption: impact of water-dip disinfestation technology*. Clin Infect Dis, 2003. **37**(12): p. 1585-90.

55. James, L., et al., *Emergency care physicians' knowledge, attitudes, and practices related to surveillance for foodborne disease in the United States*. Clin Infect Dis, 2008. **46**(8): p. 1264-70.
56. Hedberg, C.W., et al., *Timeliness of enteric disease surveillance in 6 US States*. Emerging Infectious Diseases, 2008. **14**(2): p. 311-313.
57. Centers for Disease Prevention and Control, *Salmonella Outbreak Investigations: Timeline for Reporting Cases*. 2008.
58. Hoffman, R.E., et al., *Capacity of state and territorial health agencies to prevent foodborne illness*. Emerg Infect Dis, 2005. **11**(1): p. 11-6.
59. Hopkins, R.S., *Design and operation of state and local infectious disease surveillance systems*. Journal of Public Health Management and Practice, 2005. **11**(3): p. 184-190.
60. Efron, B. and R. Tibshirani, *An introduction to the bootstrap*. Monographs on statistics and applied probability ; 57. 1993, New York: Chapman & Hall. xvi, 436 p.
61. Efron, B., *Nonparametric Estimates of Standard Error - the Jackknife, the Bootstrap and Other Methods*. Biometrika, 1981. **68**(3): p. 589-599.
62. Efron, B., *Model Selection and the Bootstrap*. Mathematical Social Sciences, 1983. **5**(2): p. 236-236.
63. Efron, B. and G. Gong, *A Leisurely Look at the Bootstrap, the Jackknife, and Cross-Validation*. American Statistician, 1983. **37**(1): p. 36-48.
64. Efron, B., *Bootstrap Confidence-Intervals for a Class of Parametric Problems*. Biometrika, 1985. **72**(1): p. 45-58.
65. Efron, B., *Jackknife, Bootstrap and Other Resampling Methods in Regression-Analysis - Discussion*. Annals of Statistics, 1986. **14**(4): p. 1301-1304.
66. Steyerberg, E.W., et al., *Internal validation of predictive models: Efficiency of some procedures for logistic regression analysis*. Journal of Clinical Epidemiology, 2001. **54**(8): p. 774-781.
67. Lynch, M., et al., *Surveillance for Foodborne-Disease Outbreaks, United States, 1998-2002, CDC Surveillance Summaries*. MMWR, 2006 **55**(SS-10): p. 1-34.
68. Sobel, J., et al., *The pandemic of Salmonella Enteritidis phage type 4 reaches Utah: a complex investigation confirms the need for continuing rigorous control measures*. Epidemiology and Infection, 2000. **125**(1): p. 1-8.
69. Kimura, A.C., et al., *Chicken consumption is a newly identified risk factor for sporadic Salmonella enterica serotype enteritidis infections in the United States: A case-control study in FoodNet sites*. Clinical Infectious Diseases, 2004. **38**: p. S244-S252.
70. Jones, T.F. and F.J. Angulo, *Eating in restaurants: A risk factor for foodborne disease?* Clinical Infectious Diseases, 2006. **43**(10): p. 1324-1328.
71. Gerner-Smidt, P., et al., *Molecular surveillance of Shiga toxin-producing Escherichia coli O157 by PulseNet USA*. Journal of Food Protection, 2005. **68**(9): p. 1926-1931.
72. Sandt, C.H., et al., *The key role of pulsed-field gel electrophoresis in investigation of a large multiserotype and multistate food-borne outbreak of Salmonella*

- infections centered in Pennsylvania.* Journal of Clinical Microbiology, 2006. **44**(9): p. 3208-3212.
73. Kaplan, J.E., et al., *The frequency of a Norwalk-like pattern of illness in outbreaks of acute gastroenteritis.* Am J Public Health, 1982. **72**(12): p. 1329-32.
 74. Sobel, J., et al., *Investigation of multistate foodborne disease outbreaks.* Public Health Reports, 2002. **117**(1): p. 8-19.
 75. Hutwagner, L.C., et al., *Using laboratory-based surveillance data for prevention: An algorithm for detecting Salmonella outbreaks.* Emerging Infectious Diseases, 1997. **3**(3): p. 395-400.
 76. Mahon, B.E., et al., *An international outbreak of Salmonella infections caused by alfalfa sprouts grown from contaminated seeds.* Journal of Infectious Diseases, 1997. **175**(4): p. 876-882.
 77. Sivapalasingam, S., et al., *A multistate outbreak of Salmonella enterica serotype Newport infection linked to mango consumption: Impact of water-dip disinfection technology.* Clinical Infectious Diseases, 2003. **37**(12): p. 1585-1590.
 78. Li, J., et al., *Evaluation of a Statewide Foodborne Illness Complaint Surveillance System, Minnesota, 2000-2006.* J. Food Prot., 2010. **In Press.**
 79. Minnesota Department of Health, *Salmonellosis, 2007.* 2007.
 80. Blanton, L.H., et al., *Molecular and epidemiologic trends of caliciviruses associated with outbreaks of acute gastroenteritis in the United States, 2000-2004.* Journal of Infectious Diseases, 2006. **193**(3): p. 413-421.
 81. Rogerson, P.A. and I. Yamada, *Approaches to syndromic surveillance when data consist of small regional counts.* MMWR Morb Mortal Wkly Rep, 2004. **53 Suppl**: p. 79-85.
 82. Lawson, A. and K. Kleinman, *Spatial and syndromic surveillance for public health.* 2005, Chichester ; Hoboken, NJ: Wiley. xiv, 269 p.
 83. Miller, B., et al., *Syndromic surveillance for influenzalike illness in ambulatory care network.* Emerg Infect Dis, 2004. **10**(10): p. 1806-11.
 84. Ploberger, W. and W. Kramer, *The Cusum Test with Ols Residuals.* Econometrica, 1992. **60**(2): p. 271-285.
 85. Hawkins, D.M. and D.H. Olwell, *Cumulative sum charts and charting for quality improvement.* Statistics for engineering and physical science. 1998, New York: Springer. xvi, 247 p.
 86. Minnesota Department of Health, *Minnesota Department of Health 2007 Gastroenteritis Outbreak Summary.* 2007.
 87. Buehler, J.W., et al., *Syndromic surveillance and bioterrorism-related epidemics.* Emerg Infect Dis, 2003. **9**(10): p. 1197-204.
 88. Kleinman, K., et al., *Evaluating spatial surveillance: detection of known outbreaks in real data.* Stat Med, 2006. **25**(5): p. 755-69.
 89. IOM (Institute of Medicine), *Review of Food and Drug Administration's Role in Ensuring Safe Food.* 2010, The National Academies Press: Washington, DC.
 90. National Association of County and City Health Officials, *2008 National Profile of Local Health Departments.* 2009, National Association of County and City Health Officials: Washington D.C.

91. Dominguez, A., et al., *Differential features of foodborne gastroenteritis outbreaks of known and unknown etiology*. J Food Prot, 2009. **72**(9): p. 1958-62.
92. Centers for Disease Control and Prevention, *Publication of updated guidelines for evaluating public health surveillance systems (Reprinted from MMWR, vol 50. pg 646, 2001)*. Jama-Journal of the American Medical Association, 2001. **286**(12): p. 1446-1446.
93. Schlundt, J., *New directions in foodborne disease prevention*. International Journal of Food Microbiology, 2002. **78**(1-2): p. 3-17.

Appendices

Appendix A

FOODBORNE ILLNESS COMPLAINT FORM
Foodborne Illness Report
Minnesota Department of Health
Phone: (612) 676-5414 Fax: (612) 676-5730

Stool kit delivered G
Daily G

Complaint date: ___/___/___ Tennessen: Q Reporter:

Agency: _____ Phone: _____ Fax: _____

First Name: _____ Last Name: _____ Age: _____ Q
Female Male

Address _____ Zip: _____

Day phone: (_____) _____ Evening phone:
(_____) _____

Occupation: _____ Daycare exposure: Yes
No

Illness History: Illness onset: ___/___/___ Time: _____ Illness Recovery Date: ___/___/___

Time: _____

Vomiting Y N Onset date: ___/___/___ Time: _____ Vomiting recovery date: ___/___/___

Time: _____

Diarrhea Y N Onset date: ___/___/___ Time: _____ Diarrhea recovery date: ___/___/___

Time: _____ Number of stools per 24 hour period: _____ Cramps Y N Fever Y N

temp: _____ Bloody stools Y N Other symptoms:

Called healthcare provider: Y N Visited provider: Y N Please circle Office / ER Date of visit:
___/___/___

Provider requested stool sample: Y N Date stool submitted: ___/___/___ Result: _____

Hospitalized: Y N

Food History:

If only one person is ill; complete entire four day food history.

If ill persons live in the same household complete entire four day food history.

If more than one person is ill and they live in different households, then record only the common meals.

Date of Illness Onset: ___/___/___

Meal Time Foods and Drinks Consumed and Location (including home)

Brk: _____

Lun: _____

Sup: _____

Oth: _____

Day Prior to Illness Onset: ___/___/___

Meal Time Foods and Drinks Consumed and Location (including home)

Brk: _____

Lun: _____

Sup: _____

Oth: _____

Two Days Prior to Illness Onset: ___/___/___

Caller=s

name: _____

Meal Time Foods and Drinks Consumed and Location (including home)

Brk: _____

Lun: _____

Sup: _____

Oth: _____

Three Days Prior to Illness Onset of Illness: ___/___/___

Meal Time Foods and Drinks Consumed and Location (including home)

Brk: _____

Lun: _____

Sup: _____

Oth: _____

Establishment or Product Complainant Suspects (for products, include brand, size, flavor, UPC, purchase date & location)

Number of persons exposed: _____ Number ill: _____

Did complainant
call the establishment:
Yes No

History of others Ill:

First name: _____ **Last name:** _____

Age: _____

Address: _____ **Phone:** _____

Illness onset date: ___/___/___ Onset time: _____ Recovery date: ___/___/___ Recovery time: _____

Vomiting Y N Onset date: ___/___/___ Time: _____ Vomiting recovery date: ___/___/___
Time: _____

Diarrhea Y N Onset date: ___/___/___ Time: _____ Diarrhea recovery date: ___/___/___
Time: _____ Number of stools per 24 hour period: _____ Cramps Y N Fever Y N

temp: _____ Bloody stools Y N

Other

symptoms: _____

Foods eaten at common event:

Agencies Notified Q MDH-EHS Q MDH-District Office Q MN Dept of Ag Q FDA
Q USDA
Q Local Agencies:

Comments

Complainant Expectations: Q Follow-up restaurants/establishments requested **Or** • MDA Follow-up

- Complaint to be logged in database only

MDH Use Only: Stool collected: ___/___/___ Received at MDH: ME I M

Results: Calicivirus O157 *Shig* *Salm* *Campy* *Yersinia*
Other _____ Negative

Notified case: ___/___/___ Notified local agency: ___/___/___

Appendix B

Foodborne Illness Complaint Surveillance Survey

1. Introduction

* 1. What is the name of your agency (i.e. Hennepin County Health Department)?

* 2. Does your agency have a consumer complaint surveillance system to allow the public to report foodborne illnesses they suspect were caused by particular products, events, or food establishments?

Yes

No

2. Complaint System Questions

1. Please indicate the type of information collected by the complaint surveillance system (choose all that apply)?

- Contact information of caller
- Information on caller (including)
- Suspected food establishment or product
- Symptom information
- Details of illness
- Ill caller sought health care
- Food history (e.g. meals eaten in 4 days prior to onset of illness)
- Biomedical samples collected
- Name of ill group
- Number exposed

2. Please comment on other information that is collected by the complaint system that was not addressed in the previous question.

3. Can we ask someone at your agency to participate in a more detailed follow up survey on your surveillance system?

Yes

No

Foodborne Illness Complaint Surveillance Survey

4. Please indicate a name and e-mail of the individual that should be contacted follow up survey.

Name
E-mail

3. No Complaint System

1. Please indicate the reason(s) why your agency doesn't have a consumer complaint system (check all that apply)?

- Complaints are taken by the state health department or other state agency
- Complaints are taken by another agency besides the state
- Complaint surveillance is not effective in detecting foodborne illness outbreaks
- Lack of personnel
- Lack of resources

2. Please comment on other reasons your agency doesn't use a complaint based surveillance system.

3. Would your agency use a complaint system if a no cost online database was to save and track complaint information?

- Yes
- No

4. End

1. Thank you for your time in completing this survey. If you would like to find out results of this survey please enter your e-mail below.

Appendix C

Second Survey

Agency Demographics / General Info

1. State
2. City
3. Agency Name
4. Number of confirmed outbreaks last year
5. Number found through complaint surveillance

General System Information

6. Approximately how many complaint calls were received last year?
7. What does your agency consider a complaint call?
8. How does the public report a complaint (check all that apply)?
 - a. Call the health department
 - b. E-mail
 - c. Online (webbased) reporting form
 - d. Voicemail
9. Is there a standard set of questions asked of each caller?
 - a. Yes
 - b. No
10. What information is collected from each caller (select all that apply)?
 - a. Caller contact information
 - b. Demographic information (i.e. age, gender)
 - c. Symptom information
 - d. Onset illness
 - e. Number ill
 - f. Number exposed
 - g. Suspected establishment/product that caused illness
 - h. Food history (how long)

Storage and Use of Data

11. How is the collected data stored?
 - a. Paper copies (skip to
 - b. Electronic database
12. If stored as an electronic database, what year does the database go back to?
13. Are complaints examined for common exposures?
14. Approximately what percentage of complaints are investigated?

- a. 0%
- b. 0-25
- c. 25-50
- d. 50-75
- e. 75-100

15. Who investigates calls? (check all that apply)

- a. Not investigated
- b. Sanitarian inspects restaurant
- c. Health officials follow up with calls that are suspected to be part of an outbreak

16. Are stool samples collected from complainants?

Data Sharing

17. Is complaint information reported to the state health department?

18. Is complaint information shared with neighboring health jurisdictions?