The Interagency Food Safety Analytics Collaboration (IFSAC) Webinar:
"Are Outbreak Illnesses Representative of Sporadic Illnesses?"
Presented on January 10, 2014

CLOSED CAPTIONING TRANSCRIPT

A recording of the webinar (audio and visual) can be found at this URL:
https://collaboration.fda.gov/p5djja1c2xe/

Speaker	Agency	Topic
Cary C. Parker, MPH (moderator)	U.S. Food and Drug Administration (FDA)	Welcome
David P. Goldman, MD, MPH (presenter)	USDA's Food Safety and Inspection Service (FSIS)	Introduction
Eric D. Ebel, DVM, MS, DACVPM(Epi), ASA/CERA & **Michael S. Williams**, PhD (co-presenters)	USDA's Food Safety and Inspection Service (FSIS)	IFSAC's outbreak and sporadic illness attribution project
David P. Goldman, MD, MPH (presenter)	USDA's Food Safety and Inspection Service (FSIS)	Closing Remarks
Michael C. Bazaco, PhD, MS (moderator)	U.S. Food and Drug Administration (FDA)	Q & A Session – Open to all attendees

Slide 1
<<CCP>> Good afternoon. My name is Cary Parker from FDA's Center for Food Safety and Applied Nutrition and with me today is Michael Bazaco. We will be serving as moderators for today's webinar. Our webinar entitled, "Are Outbreak Illnesses Representative of Sporadic Illnesses?" is sponsored by the Interagency Food Safety Analytics Collaboration (or IFSAC). IFSAC is a collaboration between the Food and Drug Administration (or FDA), the Centers for Disease Control and Prevention (or CDC) and the US Department of Agriculture, Food Safety and Inspection Service (or FSIS). This tri-agency analytic collaboration is currently focused on projects related to foodborne illness source attribution.

Today's presenters will provide more background information regarding the Collaboration and more details on one of our specific projects. Please note that there will be time at the end of the webinar for questions. Please type your questions in the "Q&A Box" on your screen and they will be answered at the end of the presentation in the order they were received. I'd also like to mention that the entire webinar will be recorded, both audio and visual, and the recording will be posted online in the near future. If you experience problems with the Adobe Connect software, please submit your technical issue in the "Q&A Box" and someone will assist you.

Slide 2
And now I'd like to introduce our first presenter - Dr. David Goldman who will provide more details about why IFSAC was formed and what IFSAC's goals and activities include. Dr. Goldman is currently the Assistant Administrator for the Office of Public Health Science at the Food Safety and Inspection Service (FSIS). He is a family practice and preventive medicine/public health physician, and a member of the Commissioned Corps of the U.S. Public Health Service, assigned to FSIS since February 2002. Dr. Goldman?

<<DPG>> Thank you, Cary. And welcome to all of you. We were very pleased with the interest in this particular webinar—we had over 800 registrants for the webinar and so again we are very pleased with your interest in this particular topic. What we would like given significant interest and registration is to cover some background information about IFSAC for those who have not participated in public meetings about IFSAC.

Slide 3
So, IFSAC is an inter-agency collaboration and it builds on a history of having worked together on source attribution over a number of years. –IFSAC was conceived about four years ago. It applies advances in source attribution methods, it leverages knowledge, expertise and data among the agencies, it builds an efficient structure for getting projects completed and it prioritizes communications and stakeholder input; and this webinar is an example of that. I'll cover each of these points in a little more detail. So, IFSAC applies whatever advances exist in source attribution methods and here listed are a few of the methods that have been or are being used in IFSAC projects currently.

Slide 4
We recently improved food categorization and that was the topic of the first webinar that we hosted back in June. We have undertaken a statistical analysis of data from foodborne outbreak surveillance from the CDC's database. We have undertaken a hybrid analysis using surveillance data from both outbreaks and sporadic cases and case control data in particular. We have adapted a modeling approach to attribution that was developed in Denmark using U.S. data. And importantly, we have taken up the issue of uncertainty by providing estimates of uncertainty where possible in the projects that we have undertaken here.

Finally, again, we've expanded the data sources we have used; not only human data sources, but also food contamination data in the IFSAC projects.

Slide 5
IFSAC takes advantage of a shared environment to develop methodology and conduct analyses. We have a very active SharePoint site that is available to all levels of those who are working on IFSAC projects from the steering committee on down to the project teams. We applied data from all sources as was mentioned just a minute ago. Importantly, the goal is to share results from analysis, their interpretation, and their use so as to enhance policy decisions from any of the three agencies involved.

Slide 6
IFSAC does consist of this structure (on slide), which I will cover briefly here. There's the steering committee comprised of two members from each of the three agencies who are able to make decisions about commitment of resources for various projects. We do rotate the chair of the steering committee among the three agencies and importantly the steering committee is intimately involved in assessment and approval of initial project proposals, as well as overseeing the progress of those projects. There's an extensive technical work group comprised of members of each of the three agencies and it is a dedicated group of agency experts, analysts, statisticians, food safety experts of various disciplines who understand the needs of their agency and who develop proposals and plans for possible IFSAC projects. They (technical workgroup) also help to coordinate IFSAC activities within each of the three agencies. Finally, and importantly, we have project teams which are comprised both of technical work group members as well as other experts from the three agencies who were pulled in to assist with specific projects.

Slide 7
Now, back to communications and stakeholder input. Here is a partial list of public outreach efforts made on behalf of IFSAC. You'll notice that it starts on this slide (Slide 6) with a series of public meetings dedicated to the issue of metrics in 2010, those were chiefly sponsored by the FDA, but also included participants from CDC and FSIS. And I won't got through each of these in detail, but I will point out that we did hold a public meeting in January of 2012 that was very heavily attended with about 250 attendees at that meeting in person and also I mentioned earlier we did host our first webinar in June of 2013 and the topic of that webinar was improving the categorization of foods implicated in outbreaks.

We have also provided periodic stakeholder updates through other means. Upcoming you will see we do intend to dedicate an IFSAC webpage. Currently you can find IFSAC information on the CDC's homepage, but we intend to establish an independent IFSAC webpage that each of the three Agencies will link to-- it may also be hosted at CDC, but it will be dedicated to IFSAC. We expect that it will be stood up within the next three or four months. Finally,

you will see we have pledged to do periodic public meetings when we have significant output to present to the public. We expect that within 12 months or so we will be able to have another public meeting in which we will present some of our significant projects.

Slide 8
The IFSAC webinars have turned out to be a low-cost and easily accessible mode of communication with stakeholders. We are able to expeditiously share project updates and results with you, our stakeholders, before publication in peer-reviewed journals, which some of you may know takes some time to do. We plan on two webinars a year. This is the second of the two this year and today we will talk about the topic of outbreak illnesses and whether they are representative of sporadic illnesses. I want to say in closing my introductory remarks that this webinar that you are about to hear is highly technical and there are some complex statistical analyses that you'll hear presented here. The presenters, who happen to be senior risk analysts from FSIS, will share an analysis comparing the characteristics of illnesses associated with foodborne outbreaks with those that are not linked to foodborne outbreaks, which we call sporadic illnesses. The fact is that most foodborne illnesses do occur as individual sporadic cases not otherwise connected or recognized as part of an outbreak. However, most information about food sources is derived from outbreak and outbreak investigations.

Back to the question; we wondered how similar are cases in outbreaks to those that occur in sporadic cases. We will address three key questions in this presentation. First, why it is important to compare sporadic illnesses with outbreak illnesses, some of which I just covered. Second, what does this study show, and third, how is it important to you, our stakeholders who have tuned in for this webinar. Again, this is a highly technical presentation. Because of this and the broad range of backgrounds among the participants we know have joined us today, I will come back and do a quick wrap-up of what you've heard during the technical presentation at the end and before our questions. With the large number of registrants as Cary mentioned we may not get to and we likely will not get to all of your questions, but we do thank you for your patience as we attempt to answer as many questions as possible. With that I'll turn back to Cary Parker and then we will get on with the technical presentation. Thank you.

<<CCP>> Thank you Dr. Goldman for the overview of IFSAC. One of the current projects that Dr. Goldman mentioned involved the evaluation of whether or not outbreak-associated foodborne illnesses are representative of sporadic (or non-outbreak) illnesses with respect to geographic, demographic, temporal, and clinical characteristics.

Slide 9
Our next presenters, Dr. Eric Ebel and Dr. Michael Williams will now elaborate on the progress from this project and describe next steps. Dr. Eric Ebel is a Senior Veterinary Medical Officer with the Risk Assessment and Analytics Staff of the Food Safety and Inspection Service (FSIS) and Dr. Michael Williams is a Senior Risk Analyst with the Risk Assessment and Analytics Staff of FSIS. Dr. Ebel?

<<EDE & MSW>> Thank you, Cary. The work we are presenting today comes from a Project Team under the leadership of the Interagency Food Safety Analytics Collaboration. Our work examined the question: are foodborne illnesses associated with outbreaks representative of sporadic – or non-outbreak – foodborne illnesses?

For those of us working in food safety – and using outbreak data to understand the food sources responsible for foodborne illnesses – this question has long been on our minds. This project attempts to address the question using available data.

Slide 10
In our work, we distinguish between outbreak and sporadic (or non-outbreak) foodborne illnesses. Outbreak cases are those that occur in an outbreak setting – usually a scenario wherein two or more people become ill under similar circumstances with respect to timing and/or location of exposure. Sporadic cases are best defined as cases not associated with an outbreak – in other words, they are single cases not overtly linked to other cases.

As we will discuss, outbreak cases provide insight into the most probable food source for those illnesses. But, we want to know how applicable that source information might be to sporadic cases. The degree to which outbreak cases represent non-outbreak cases can inform other source attribution questions contemplated by IFSAC. For example, the confidence we have in our conclusions will also contribute to other work that intends to describe uncertainty about source attribution estimates. But, to be clear, the work we will describe today does not estimate source attribution fractions.

Slide 11
As background, keep in mind that attributing illnesses to their sources is not trivial. We need to know that an ill person is sick because of a specific pathogen and we need to determine what food was the source of that pathogen.

A natural resource for this information is the CDC's National Foodborne Disease Outbreak Surveillance System, or FDOSS. This passive national surveillance system collects reports – from state, local, and territorial public health agencies – of foodborne outbreaks caused by enteric bacterial, viral, parasitic, and chemical agents. The FDOSS database contains investigation findings that include the most probable food source for many reported outbreaks. We can tally outbreaks (or the numbers of illnesses associated with those outbreaks) and determine the fraction of the total associated with particular food commodities.

Given that FDOSS data are amenable to estimating food source attribution directly, we can ask; why not just consider the FDOSS data as applicable to sporadic illnesses? The answer is that we'd like to, but there are some challenges. One such challenge is that outbreak cases represent just a fraction of all cases. For example, of the one million or so foodborne Salmonella illnesses we think occur each year in the United States, there are around 3000 (or just 0.3%) reported outbreak cases.

Slide 12
The Foodborne Diseases Active Surveillance Network (or FoodNet) conducts surveillance for a number of enteric infections diagnosed by laboratory testing of samples from patients. FoodNet is a collaborative program among 10 state health departments and the Federal agencies involved in IFSAC. FoodNet personnel located at state health departments regularly contact the clinical laboratories in the sites listed here to get reports of infections diagnosed in residents of these areas. The surveillance area includes 15% of the United States population (or about 48 million persons). The FoodNet surveillance network finds sporadic illnesses mostly. Those cases do not point to a most probable food source, although a few case-control studies have been conducted in the past to identify foods more commonly associated with cases than non-cases.

Slide 13
We want to know if the characteristics of outbreak cases are representative of non-outbreak or sporadic cases. Specifically, we would love to know if source attribution estimates are comparable between outbreak and non-outbreak cases. Unfortunately, as we described before, because sporadic illnesses captured in FoodNet do not include information about exposures, there isn't a way to directly estimate source attribution using these data alone.

So, we do not have a way to assess the representativeness of outbreak-derived attribution estimates to non-outbreak cases directly. Nevertheless, we decided to assess these two populations for other characteristics to see how comparable they were.

Slide 14
This is the hypothesis that we decided to test in our analysis. We reasoned that, if outbreak and non-outbreak cases are similar with respect to characteristics that we CAN measure, then there is no justification for rejecting their similarity with respect to food source attribution. Nevertheless, if the available data suggest that outbreak and non-outbreak cases are different with respect to these characteristics, then the existing evidence does not support their similarity and this makes the assumption about similar attributions less plausible.

Slide 15
This is an overview of the project. We compared characteristics between outbreak and non-outbreak cases. These characteristics pertained to geographic, demographic, temporal and clinical factors that were available to us in a database.

We did this analysis for the four pathogens shown here. For federal food safety officials, these are the pathogens of highest concern.

As we just discussed, if the data supports similarity between outbreak and non-outbreak cases, then we cannot reject the plausibility that source attribution estimates might also be similar. This line of reasoning is similar to any circumstance where we sample a part of a population and wish to apply an estimate from this sample to the entire population. If the sample is reasonably representative of the general population, then the sample estimates should be reasonably applicable to the general population.

Slide 16
To conduct this analysis, we used FoodNet data because these data provide information on both sporadic and outbreak cases. Although FoodNet is designed to actively identify sporadic illnesses, some of the illnesses in FoodNet are eventually identified as having been associated with outbreaks.

We used FoodNet data from 2004 through 2011 because the surveillance system maintained the same 10 states during this period – and we conducted the analysis before more recent data were finalized.

As this table shows, a very small percent of FoodNet Campylobacter cases are outbreak cases, but almost 20% of STEC cases are outbreak cases. Outbreak cases comprise about 5% of Listeria and Salmonella FoodNet cases.

Slide 17
These are the predictor variables we included in the study. To capture geographic differences, we used the FoodNet State where cases occurred. For temporal effects, we used the year and season of each case. For demographic effects, we used the age and gender of cases. As a measure of clinical severity, we used a data field that indicated whether a case resulted in hospitalization.

Slide 18
In our minds, these predictors fell into two different classes. The STATE, YEAR and SEASON predictors were considered structural, or surveillance, factors. While outbreak and non-outbreak cases might have different occurrences across these factors – and these differences might signal differences in exposures between these types of cases – we would not consider such differences as epidemiologically meaningful. For example, variability we might see across states or years may be influenced by administrative and resource factors associated with running a surveillance system rather than some fundamental difference in the biology of outbreak and sporadic cases. Nevertheless, these structural factors need to be considered in the statistical analysis. Also, current approaches to estimating food source attribution typically aggregate data across these surveillance factors – so they tend to be averaged out in the process.

The other predictors were classified as Case factors. If we observed differences between outbreak and non-outbreak cases with respect to age, gender or hospitalization, then such differences seem meaningful epidemiologically, especially because we are controlling for structural factors. For example, if we find – for a particular pathogen – that outbreak cases are more or less frequently severe than non-outbreak cases, then the notion that these cases have similar exposure and clinical patterns is difficult to argue. These types of differences will not “average out” when we aggregate data for estimating food source attribution and could result in biased estimates.

Slide 19
In our analysis, we needed to define seasons and age categories. To do this, CDC defined quintiles for each pathogen and factor such that 20% of the data resided within each interval. In this chart, the quintiles are represented by the 5 different colored bars. The seasonal quintiles were similar across pathogens except, for STEC, the first season extended from January through the end of May.

The age quintiles differed substantially across pathogens. However, for example, we needed to define the first quintile for Listeria as those cases from 0 to 38 years of age in order to capture 20% of the data for that pathogen. In contrast, the first quintile for Salmonella only extends to 3 years of age; this suggests that 20% of Salmonella illnesses occur in that age category. As we'll see later, this is an important age class in our analysis.
Another interesting pattern for Listeria is the relatively narrow quintile ranges for persons over 60 to 80 years of age. This reflects the large number of cases among older people for this pathogen.

Slide 20
To complete our analysis of these predictors, we used a two-step approach. First, we conducted a Random Forest analysis to essentially "triage" the factors as to their relative importance. Random Forest analysis is a data classification algorithm. In our analysis, the Random Forest algorithm searches among possible combinations of predictors to find those combinations that best explain the occurrence of outbreak and non-outbreak cases in the FoodNet data. We used the Random Forest analysis because it is better than regression modeling for dealing with interactions between predictors and for handling missing observations. We eliminated predictors from further consideration if the Random Forest analysis showed their influence on predicting outbreak and non-outbreak cases was insubstantial. Factors that were not eliminated were carried into the next step.

Our next step was logistic regression modeling. Regression modeling allows us to easily interpret the importance of individual predictors. It also allows us to conduct model selection using metrics that are generally understood among a broader audience. In this step, we collapsed the data by factors (and/or combinations of factors) such that the dependent variable was the fraction of outbreak cases among all cases for a specific combination of predictors. For example, one line of the Salmonella dataset might refer to all cases from Tennessee that occurred between July and August of 2006, and that were classified as 14 – 34 years olds. This same line's dependent variable would be the fraction of cases matching these predictors that were outbreak cases.

The logistic regression step gave us a formal method for determining what predictors best explained being an outbreak case. If the answer was, "no factors explain a difference between outbreaks and non-outbreaks," then we would not reject our null hypothesis.

Slide 21
[Results – transition slide]

Slide 22
In the Random Forest analysis, we examined each pathogen independently. We started each analysis assuming all six of our predictors might be relevant in explaining differences in the occurrence of outbreak and non-outbreak cases. The Random Forest technique searches among classification trees to find those combinations of predictors that perform best in explaining these differences.

Using data that are set aside for validation, misclassification statistics generated by the Random Forest software quantify the accuracy of alternative classification trees. For all four pathogens, models that included the Gender and Hospitalization predictors performed no better than models that only included the other four predictors. Therefore, this analysis determined that the Gender and Hospitalization predictors were not important; and these predictors were dropped for the logistic modeling step.

Slide 23
This graph illustrates why Gender and Hospitalization were not important coming out of the Random Forest analysis. The y-axis on the graph represents the fraction of cases that were outbreak cases for each of the

categories on the x-axis. Across the four pathogens, we see little difference in the fraction of cases that are outbreak cases for either the Gender or the Hospitalization predictors. The Random Forest analysis tells us that the small differences we do see were not important.

This graph also reminds us of the different fractions of outbreak cases for each pathogen. Recall that this fraction was about 0.4% for Campylobacter, just under 20% for STEC and around 5% for Listeria and Salmonella.

Slide 24
The logistic regression modeling was conducted for all four pathogens using the remaining predictors of STATE, YEAR, SEASON and AGE. We used a step-wise fitting algorithm to examine alternative models comprising these predictors and their possible combinations.

Because regression modeling techniques always show a better fit as more predictors are added to a model, we used a Bayesian Information Criterion (or, BIC) statistic to select the preferred model. The BIC provides a measure of the statistical model's fit with a penalty for each additional predictor. Practically, we selected the model with the minimum BIC value.

The results of the logistic modeling are shown here. All the models included the State predictor. The STEC and Listeria models also included the year predictor. Only the Salmonella model included all four predictors and some of their interactions.

Because of its complexity, we'll look at the Salmonella results a little more carefully.

Slide 25
These graphs simply illustrate the BIC values for alternative models for STEC and Salmonella. Recall that the smallest BIC value indicates the best balance between explanatory power and model complexity. The STEC graph (on top) illustrates that the smallest BIC value occurs for the model with the 10 State parameters and the 8 year parameters. The Salmonella graph (at the bottom) shows that the minimum BIC value does not occur until we incorporate more of the predictors and their interactions; so the best model has just over 150 parameters.

Slide 26
To illustrate the effect of the STATE predictor, this graph shows the fraction of all cases in the different FoodNet States that are outbreak cases. The grey bars show the outbreak fraction for STEC illnesses while the blue bars show the outbreak fraction for Salmonella illnesses.

Reading off the left y-axis, STEC outbreak fractions range from 38% in Colorado to 2% in California. This range illustrates why the STATE predictor is associated significantly with the probability of being an outbreak case. Reading off the right y-axis, we note that the pattern of Salmonella outbreak fractions across states is not exactly the same as we see for STEC illnesses. Nevertheless, California also has the smallest fraction of Salmonella outbreak cases and this fraction ranges from over 20% in Oregon to the 2.5% in California.

Slide 27
To appreciate how outbreak fractions might change across states and years for Salmonella, we look at a couple of low outbreak fraction states in the left graphic here and a couple of high outbreak fraction states in the right graphic. In both graphs, the left y-axis measures the number of non-outbreak cases, while the right y-axis measures the number of outbreak cases. The solid lines relate to non-outbreak cases while the broken lines of the same color relate to outbreak cases.

Generally, in both graphs, we can appreciate that non-outbreak cases are relatively stable across years, but outbreak cases are more variable. Interestingly, what seems to distinguish the left graph from the right graph is not the numbers of outbreak cases (which have similar scales ranging from 0 to ~120 or 140 cases per year), but, instead, the dramatic differences in non-outbreak cases in the left graph as compared to the right. In other words, what seems to explain the low outbreak fraction states is that those states have many more non-outbreak cases

than the high outbreak fraction states. It is speculative to explain why these differences exist, but resourcing and administration of the programs may play a role.

Slide 28
In the best Salmonella model, we found significant interactions between some of the predictors. To illustrate these, we developed interaction profiles. In the next slide I'll show these profiles.

One example of an interaction was illustrated in the previous slide, where we showed that outbreak counts varied across years and states. Such interactions may just be idiosyncrasies associated with a surveillance system that operates across 10 different public health jurisdictions, but it cannot be ruled out that these differences might reflect different food exposure pathways between outbreak and non-outbreak cases for different states in different years.

As we explain here, crossed lines in the interaction profile graphics imply interactions between two predictors while parallel line structures suggest no interactions in the predictors. Because there were no interactions found for the STEC, Campylobacter and Listeria analyses, we will just show the profile for Salmonella.

Slide 29
The four predictors are listed in the diagonals of this graph. The y-axes measure the fraction of cases that are outbreak cases. The x-axes represent the FoodNet states, surveillance years or the quintiles for the SEASON and AGE predictors.

If we look at the graph in the upper left, just below the state box and just to the left of the Year box, we see how outbreak fraction varies across the 10 FoodNet States, with each line representing one of the eight surveillance years. The fact that some of these lines cross indicates that outbreak fraction variability across states is not constant for every year.

The graph that is up and to the right of this graph shows the same relationship, but in reverse so that the x-axis is surveillance year and the lines represent different states. The interpretation is the same for both graphs, however.

The other graphs that illustrate crossed lines are those relating YEAR and SEASON. In contrast, there is no indication of interaction of AGE with any of the other three predictors.

Slide 30
We looked at the distributions of ages among non-outbreak and outbreak cases for Salmonella. Because AGE was a case characteristic, its significance in the Salmonella model warrants closer examination.

What we see in this graph is that a large share of the non-outbreak cases occur in the 0-3 year-old age range, while a much smaller share of outbreak cases occur in this age range. We can only speculate an explanation of this dramatic difference, but its existence suggests that food source attribution fractions estimated from outbreak data may not be applicable to the youngest non-outbreak cases. There are some work-arounds to this issue that we can explore. But, we think it is important to recognize that essentially 20% of Salmonella sporadic illnesses in FoodNet occur among these youngest ages. Based on our results, it is not plausible to assume that the exposures responsible for these cases are similar to the exposures gleaned by looking at all the outbreak information.

Slide 31
Season was a significant predictor of outbreak cases in the Salmonella model. This graph summarizes the monthly occurrences of outbreak and non-outbreak cases in the FoodNet data. It shows that the peak frequency of non-outbreak illnesses occurs around August while the peak for outbreak cases is around June. In this view, the differences do not appear substantial nor especially explanatory.

Slide 32
So, that is our analysis. Generally, we find that the case factors we were able to evaluate from FoodNet for all four pathogens suggest that outbreak and non-outbreak cases look similar. Such similarities do not reject the idea that outbreak cases are reasonably representative of non-outbreak cases and that food source attribution estimates generated from outbreak investigations may be applicable to the population of all illnesses associated with a particular pathogen.

As we've described, this general statement is not applicable to Salmonella illnesses among the youngest quintile. Nevertheless, there are methods that could be developed to allow outbreak-based food source attribution estimates to be applied to the majority of Salmonella cases.

All four pathogen models included at least one significant surveillance factor. It is not surprising that outbreak case occurrence among FoodNet cases would vary across States and/or years. Many analysts aggregate outbreak investigation data across all States and multiple years when estimating source attribution. These findings support aggregating outbreak data in this fashion because there are apparent differences in the occurrence of outbreak and non-outbreak cases across space and time. Aggregating outbreak data will help to average out these effects when estimating attribution fractions.

Slide 33
In summary, our work can only indirectly assess the applicability of outbreak-based attribution estimates to non-outbreak populations. We do not have data to make direct comparisons. Instead, we assessed how representative outbreak cases were of non-outbreak cases from the national FoodNet surveillance system.

We found little evidence to reject that case characteristics were similar between outbreak and non-outbreak cases for Campylobacter, STEC and Listeria. We should mention the fact that outbreak cases only constituted 0.4% of all cases in the Campylobacter FoodNet data; this makes the power of our conclusions weak for this pathogen.
For Salmonella, we found differences in the age distributions between outbreak and non-outbreak cases suggesting that we should be careful in how we apply outbreak-based food source attribution estimates for this pathogen. For this pathogen it is possible that the exposure rates differ between outbreak and sporadic cases, in which case outbreak data may not be representative of sporadic illnesses.

Otherwise, the fact that gender and hospitalization status are not different between outbreak and non-outbreak cases for all four pathogens suggests we have found little evidence to reject the value of outbreak-based attribution estimates for application to the larger population of non-outbreak or sporadic illnesses.

Slide 34
And with that we'd like to conclude and thank our collaborators that are listed here for this project.

<<DPG>> Thank you Dr. Ebel and Dr. Williams for that very interesting and detailed presentation. Before we go to the questions that you might have based on what you just heard, let me spend just spend a minute or two highlighting some of the important findings, really punctuating some of what you just heard. First, why was it important for us to consider sporadic illnesses when estimating source attribution? Again, we said outbreaks contribute only a small proportion of lab confirmed foodborne illnesses and the results of attribution analyses using outbreak data are often applied to sporadic illnesses. So, it was important for us to know whether these outbreak-associated cases were representative of sporadic illnesses. You also heard it's impossible usually to know for a sporadic case what food or other exposure was the source of that illness without conducting a special study. It was important for us to know if sporadic illnesses and outbreak illnesses are similar so we can better understand source attribution results that are derived from analyses of outbreak data. So, what did this study show? Again to reiterate, in general, outbreak cases closely resemble sporadic cases for three of the four pathogens that were examined; Campylobacter, Listeria and E. coli O157:H7. Keep in mind that for Campylobacter sporadic illnesses, although they are similar to outbreak illnesses, there were a very small number of outbreak illnesses available for comparison, which limits some of the conclusions that can be drawn. For Salmonella, outbreak cases resemble sporadic cases in older children and adults, but there is an important difference when you look at infants and

young children aged three and less. This study reminds us that Salmonella infections in the very young can come from other sources in the environment besides the foods that cause outbreaks.

Finally, why is this issue important to you, our stakeholders, to regulators, and to the industry? It is clearly important. The findings of this webinar are important for future IFSAC projects, which often focus on outbreak associated cases. It does mean that using outbreak data for attribution in the general population is reasonable for the most part. IFSAC will continue to study data related to foodborne illness as new information becomes available and we will continue to refine our methods for analyzing illness cases, whether sporadic or outbreak-associated, in order to come up with better estimates of attribution. We will now turn to any questions you may have and please use the Q&A box to log in your questions and we will take your questions as they come in.

Slide 35

<<MCB>> Dr. Goldman, thank you. This concludes the presentation portion of the webinar. If you have not entered your questions, please do so at this time so that we may review them. If you have already submitted a question, there is no need to re-enter it. Now we will begin the question and answers…the first question:

Question & Answer Session

Q1: "How will whole genome sequencing data be used to supplement these analyses?"

A1: This is Eric. We are not sure how that will augment or supplement this analysis but clearly the more information we have to characterize both sporadic and outbreak cases, the more detail we can apply in future analyses of this question. Nevertheless, at some level, of course, there's going to be so much specificity as to raise the question - is it meaningful, the differences we are finding? - If we get so much power into the analysis. But, that sort of analysis, and other characteristics one could imagine, data being developed for comparing outbreak and outbreak cases would certainly allow for more specific measurement of those similarities and differences.

<<MCB>> Next question.

Q2: "How do you determine if sporadic cases are foodborne or from another source such as waterborne, animal contact, etc?"

A2: In the context of this analysis, we did not make that determination in either the cases of sporadic illnesses within Foodnet or the outbreak cases within Foodnet, although there is a field within Foodnet that will give more insight into the exposure pathway for outbreak cases. Of course, we don't have that kind of information with regard to the non-outbreak cases. We did not use that field for outbreak cases in this analysis, so there certainly could've been outbreak cases in FoodNet that were not necessarily foodborne just as there might have been sporadic cases within FoodNet that were not necessarily foodborne.

<<MCB>> Thank you.

Q3: "While performing modeling techniques, how did you account for the fact that outbreak cases are not independent events? For example, many cases are related to other cases by nature of belonging to a cluster or outbreak."

A3: Yes. In a more detailed discussion of our choice of the Bayesian information criterion, that forms one of the reasons for why we believe the BIC statistic was a more appropriate statistic for evaluating fit because we did not have a way to be able to directly account in the data for the possibility of dependency between some of the outbreak cases within the data. We argued that using a BIC statistic, which would be somewhat more severe in its treatment of more complex models, was an appropriate technique for evaluating model fits.

Q4: "How would differences in health care seeking behavior affect explanation of age distributions?"

A4: It's difficult, I guess, for us to think through that question in real time. It seems clear that there are likely differences in the health care seeking behavior between youngsters and other adult aged people. We just really can't fathom how that might change the conclusions that we saw. One of the things we don't see the same kind of pattern and outbreak cases as non-outbreak cases. So we think the difference we are seeing is appears to be real rather than something that is differentially affecting the non-outbreak cases because these are all FoodNet cases in the analysis we completed.

Q5: "Did you look at individual Salmonella serotypes in your modeling?"

A5: We did do some of that analysis. We did not present it today. One of the reasons we did not incorporate it into the general analysis and report that we shared with you today is because we didn't have the capacity to do that for all four pathogens. We wanted to keep the analysis common between the four pathogens. But, we did look at that and we don't really have anything to share with you. There were similarities in the distribution amongst outbreak and non-outbreak cases. In a very general sense, the sero-type distributions that we observed in those two populations were similar but we don't have the data today to show.

<<MCB>> Are there any more questions? Here comes one.

Q6: "Do you think the analysis can be logically extended to non O157:H7 STECs?"

Q6: That's a tough question. I would assume that we would at least want to evaluate the available data, which we have not, for those pathogens so it's difficult to imagine that we would just extrapolate our findings to the non-0157 without at least evaluating the data that we do have for that group of pathogens.

<<MCB>> In the interest of time, our presenters will take your final question now. Please feel free to send an e-mail to the IFSAC e-mail which is IFSAC@fda.hhs.gov. The final question for today is…

Q7: "What would explain difference in age group between sporadic and outbreak cases of Salmonella. You suggest this could be due to non-food sources; however aren't non-food sources captured equally in sporadic and outbreak cases?"

A7: We would just be speculating. We do not have the level of detail in the data to be able to really examine that question at the level it needs to be examined. Certainly the idea that, for whatever reason, we see a higher frequency of cases in that youngest age category among non-outbreak cases suggests the possibility that there are nonfood sources or exposure pathways that are contributing to those cases. For whatever reason, those cases are not occurring in an outbreak setting. At this point, we can only speculate, but that's the line of reasoning of the questioner and it's kind of the line of reasoning that we've taken as well.

Slide 36

<<MCB>> Thank you to our presenters, Dr. Goldman, Dr. Ebel, and Dr. Williams. We appreciate your participation in today's webinar and we hope that you will join us again in the future. If you have any additional questions or suggestions for improving the logistics of your webinar experience for the next time, please send us an email at IFSAC@fda.hhs.gov. A recording of this webinar will be posted online in the near future. One final note: we will be launching an IFSAC website in Winter-Spring 2014. Please be on the lookout for an announcement soon. Again, thank you for your attendance. This concludes today's webinar session.

[Webinar concluded]

ADDENDUM:
Remaining Questions from Q & A Session with Responses

Due to the time constraints during the live webinar, we were unable to answer all questions posed by the attendees at that time. However, we wanted to make sure the remaining questions were addressed now. Below are the remaining questions (RQ) with corresponding answers.

RQ1: "Are stillborn accounted for in the 1st quintile for Listeria?"
The first quintile of reported age for the Listeria analysis included isolates from persons <1 through 38 years old, which includes Listeria isolates obtained from stillbirths. It also includes isolates obtained from pregnant women 38 years or less who had a stillborn birth outcome. Consequently, this age group accounts for more than 90% of stillborn births.

RQ2: "Very interesting new work. What are your next steps in terms of food attribution?"
Thank you for your question and for attending the webinar! IFSAC is engaged in a number of different projects to continue our efforts to improve our understanding and use of foodborne illness attribution. We are currently working on a number of projects that address different facets of foodborne illness attribution, such as the following: a project to evaluate a pathogen subtype model to better estimate the number of *Salmonella* illnesses associated with different food sources; a project to estimate foodborne illness source attribution for illnesses caused by *Salmonella*, E. coli O157:H7, Listeria, and Campylobacter; and a project to explore statistical modeling approaches to evaluate temporal trends in attribution estimates. IFSAC plans to hold a public meeting in the Fall/Winter of 2014/2015 to share results from an on-going project to develop tri-agency approved, harmonized attribution estimates. We also plan to launch an IFSAC webpage, which will be hosted by the Centers for Disease Control and Prevention (CDC), in the next several months that will provide more information about our activities.

RQ3: "Do you see food testing becoming common place with farms, produce distributors, retail stores? If not, where do you see food safety testing being conducted?"
Thank you for your question and for attending the webinar! In fact, the two federal regulatory food safety agencies--the Food and Drug Administration (FDA) and the Food Safety and Inspection Service (FSIS) within the US Department of Agriculture (USDA)--conduct food testing in a variety of settings on a variety of food products. For more information about food testing conducted by FSIS, which regulates meat, poultry, and processed egg products, please visit the following website: http://www.fsis.usda.gov/wps/portal/fsis/topics/data-collection-and-reports/microbiology. For more information on the role of testing as a verification measure in a food safety system, please visit the FDA Food Safety Modernization Act (FSMA) proposed rule on preventive controls for human food: http://www.regulations.gov/#!docketDetail;D=FDA-2011-N-0920

RQ4: "Might these findings indicate some value to changing current standard practice to partially investigating foodborne illness complaints involving just one person?"
Outbreak data are very useful for attribution studies because the source of illness can often be confirmed through an investigation. Most ill persons can't determine the source of their infection, and it is very rare that an investigation of a single illness caused by *Campylobacter*, Shiga toxin-producing *E. coli* O157, *Listeria monocytogenes* or *Salmonella* can determine the source. Many health departments partially investigate lab-confirmed illnesses involving just one person by interviewing the patient. Further investigation is generally warranted when epidemiologists identify a highly suspect source that can be tested or, more commonly, when these interviews of several patients or information about their isolates suggest a common source.