

However, while outbreak data is one of the only sources of information on the food items associated with an outbreak, it does, of course, have limitations, including the assumption that food sources in outbreaks are similar to those in sporadic illness.

While more work is certainly needed, IFSAC has done some analyses to evaluate this relationship and found that, in general, illnesses from outbreaks can be generalized to the broader population.

You can find a webinar on this analysis, conducted by Dr. Eric Ebel et al, on the IFSAC webpage and in a peer-reviewed journal article, which are also available on the journal webpage.

Slide 14: IFSAC Estimates for 2012

The work we are presenting today is an outgrowth of a report IFSAC presented at a public meeting in 2015 that many of you probably attended. In that report, we described IFSAC's new approach for estimating foodborne illness source attribution for IFSAC's four priority pathogens, which are *Salmonella*, *E. coli* O157, *Listeria monocytogenes*, and *Campylobacter*, and based on 15 years (1998-2012) of CDC outbreak data.

This approach had a number of things going for it, including using a new IFSAC-developed and published categorization scheme, that actually just—the publication was just announced the other day. It also addressed biases and adjusted for outbreak size, down-weighted the influence of older outbreaks on the final estimates and used Bayesian bootstrapping to calculate uncertainty around the estimates. If you would like more information about the analysis or data through 2012, please visit the IFSAC webpage for a link to the public meeting and the report itself.

Slide 15: Developing Estimates for 2013

Building on this report – on this approach, we developed a standard report framework to allow us to routinely update these harmonized attribution estimates as new data becomes available.

Today, we are presenting estimates with data through 2013, with some modifications. Moving forward, as we'll discussed later in the presentation, we will continue to release reports on an annual basis. Now, I'm going to pass things off to Mike Batz from the FDA to talk about our methods. Go ahead Mike.

[Recording Timestamp = 12:20](#)

Slide 16: Methods

<<MICHAEL BATZ>>: Good morning and thank you, Joanna. I'm going to talk about some of the methods we use to estimate attribution percentages. We realize that some of you attending today may not be technically inclined but we think it is important to be transparent. We're also providing more [in this presentation] than what is available in the annual estimates also being published today.

Slide 17: Our Overall Approach

As Joanna noted, our approach to estimating foodborne illness attribution for 2013 is very much the same as our approach used for 2012 estimates. The main difference is that instead of 15 years of data, we now use 16. Specifically, we utilize data on foodborne outbreaks with a single causal pathogen and which can be assigned to a single food category, from 1998 through 2013.

While most outbreak attribution efforts calculate attribution based on the numbers of reported outbreaks or the numbers of reported outbreak illness, we employ statistical models of outbreak size in

order to mitigate the influence of outliers and to incorporate epidemiological factors. This approach also uses temporal weighting to give greater influence to more recent outbreaks.

At this point, it is worth noting that what we are presenting today is not yet peer-reviewed in the scientific literature. We are close to submitting a manuscript, but it is possible that our methods may change based on this peer review, or that we will continue to advance it as we work in this area.

Slide 18: U.S. Outbreak Data

Where do we get our data on foodborne outbreaks? This data is built upon the dedicated efforts of state, local, and territorial health departments who have the primary responsibility for identifying and investigating foodborne outbreaks in the United States, and reporting them to CDC. Electronic reporting started in 1998, which is one of the reasons that's the starting year that we use.

Since 2009, this reporting has been done through a web-based platform known as NORS, or the National Outbreak Reporting System. NORS collects reports of waterborne and foodborne outbreaks, and outbreaks of enteric disease. These outbreaks can include those caused by bacterial, viral, parasitic, chemical, toxin, and unknown agents, as well as waterborne outbreaks of non-enteric disease. Data from NORS on foodborne outbreaks is [then] captured by FDOSS, or the Foodborne Disease Outbreak Surveillance System.

For each outbreak, we have information, where it is available, on the dates and locations of outbreaks, which pathogens were responsible, which foods were implicated, the number of illnesses, hospitalizations, and deaths associated with the outbreaks, as well as information on contributing factors and many other relevant variables.

Slide 19: Data Decisions

For the sake of transparency here, we think it is important to clarify which outbreaks we include in the analysis and which we don't. Let's start with three key assumptions.

First, we include outbreaks with both confirmed and suspected etiology. Etiology is a technical term that just means the causal hazard, in this case it just means the pathogen that was responsible for the outbreak. Etiology is confirmed when the pathogen or pathogens have been isolated from at least two patients or from an implicated food. By this definition, about 90% of outbreaks are confirmed with about 10% as unconfirmed, or suspected. Of those with unconfirmed etiology, about 95% of them have at least one illness where the pathogen was isolated. In our exploratory analysis, we found no significant difference between outbreaks with confirmed or suspected etiology and therefore we include those with suspected etiology in our analysis. Including those with suspected pathogens does not much change our point estimates, though it does tighten our credibility intervals a bit.

Second, and similarly, we include outbreaks where the foods implicated were either confirmed or suspected. The definition of a confirmed food vehicle is a bit different [than etiology]. In this case, a food vehicle may be considered confirmed by various types of evidence such as laboratory, epidemiological evidence, environmental assessments, or other kinds of information gathered by the investigators. For a single state outbreak, a single form of evidence is typically enough to consider a food confirmed, but for multi-state outbreak there must be two types of evidence. We do not only rely on outbreaks with confirmed food vehicles in part because of the different requirements for single state and multi-state outbreaks, in part because some of these types of evidence are more conclusive than others, and in part

because we simply don't have enough outbreaks that meet these conditions to be able to limit our analysis to them.

Lastly, it is worth noting that when we talk about the number of illnesses associated with an outbreak, we are using the total number reported, not just the number of illnesses confirmed via laboratory. For multi-state outbreaks in which data is submitted by CDC, the number of total reported outbreak illnesses is defined, essentially, as the number of lab-confirmed illnesses, whereas, for a single state outbreak, the total reported illness may include those identified by epidemiological investigations and so forth.

All three of these assumptions are in line with prior estimates of foodborne illness source attribution, such as the Painter *et al.* study from 2013 with which some of you may be familiar.

Slide 20: Data Preparation

Now, I would like to talk a bit about getting from our initial data pull to our final data set for analysis via a series of inclusion and exclusion criteria. At the top of this tree, you see that we start with over 17,000 foodborne outbreaks from 1998 to 2013. The left branch immediately below this shows the number of outbreaks excluded from further analysis, and the right branch shows what was included.

So of the 17,342 total foodborne outbreaks, 2,919 (or 16.8% of them) were due to one or more of the four specified priority pathogens. The remaining 83.2% were excluded because they were due to other pathogens. Going down the tree, we then exclude about 3% of these outbreaks due to them being associated with multiple pathogens. About 38% of the remaining outbreaks were then excluded because they did not have an identified food vehicle. Of the 1,748 outbreaks with an identified food vehicle, about 60% could be assigned to a single food category. The rest had multiple implicated foods or multiple ingredients suspected, and they were excluded from the analysis.

Here, as an aside, I do want to note that we are working on ways to utilize these 40% of outbreaks, sometimes referred to as “complex foods” outbreaks. Lastly, we excluded a few outbreaks that occurred in outlying U.S. territories, such as Puerto Rico. We end up with 1,043 outbreaks at the very bottom, which is about 37% of the single etiology outbreaks. Of these, 638 (or 60% or so) are due to *Salmonella*. Although we do have over 200 [*E. coli*] O157 outbreaks and 176 *Campylobacter* outbreaks, there are only 26 outbreaks due to *Listeria monocytogenes* in our final dataset. Moving from the type of data that we are using in our model, I would like to talk a little bit about how we assigned these to foods.

Slide 21: Food Hierarchy

This figure shows the categorization scheme as it was used for this analysis. This scheme is presented in some detail on the IFSAC website, and more information about it can be found in a peer-reviewed article titled “An Updated Scheme for Categorizing Foods Implicated in Foodborne Disease Outbreaks: A Tri-Agency Collaboration” by Richardson *et al.* published this month in Foodborne Pathogens and Disease. It has been available online for a few weeks now.

In this scheme, foods are first organized into land animals, aquatic animals, and plants, and so on into narrower categories. The full scheme includes more hierarchical levels than shown here, but for the purpose of this analysis, outbreaks are assigned to one of 22 food categories represented here by green boxes. Due to sparse data, however, we combined eight of these categories into three aggregated categories as shown in purple. Other Meat and Other Poultry were grouped into Other Meat and Poultry. Shellfish and Other Aquatic Animals were combined into Other Seafood, and Fungi, Herbs, Root-

underground, and Nuts-Seeds were combined into Other Produce. That leaves us with 17 categories. Now let's look a little bit [closer] at what's in these data.

Slide 22: Number of Reported Outbreaks by Pathogen, Food Category, and Year

There's a lot going on here so I'm just going to [walk] you through this quickly. What this figure is showing overall is the number of reported outbreaks in every year, organized by pathogens and foods. The four panels from top to bottom go *Salmonella*, *E. coli* O157, *Listeria monocytogenes*, and *Campylobacter*. Each row within each panel is a different food category organized the same as the hierarchy I showed you before. Sorry if these are too small, we just couldn't address [the size] entirely for this graphic. The color here is indicating the number of outbreaks. White indicates zero outbreaks, levels of pink go up until you get to bright red which reflects 20 or more outbreaks due to that pathogen food combination in a given year.

There are few patterns here that I think I just worth mentioning. I think clearly this reinforces the signal that we have more outbreaks for *Salmonella* than other pathogens, and it also shows that the outbreaks due to *Salmonella* are distributed across a fairly wide variety of foods. Some pathogen-food category pairs have outbreaks every year -- or many outbreaks every year. If you look at *E. coli* O157 in Beef or *Salmonella* in Eggs, you can see color all the way across. Other pathways have outbreaks routinely across the time frame but not every year. For example, *E. coli* in Fruits, or *Campylobacter* in Fish or Other Seafood products show some white spots indicating there aren't outbreaks every year, but there are generally outbreaks across the time period.

This figure also shows that some pathways have outbreaks only in the most recent years. The best example of this is that you can see there are no *Listeria monocytogenes* outbreaks associated with Turkey in the later time period -- none since 2005. Likewise, since 2008, there are number of *Listeria* outbreaks associated with Fruits, Sprouts, and Vegetable Row Crops, and none prior to that. These kinds of general characteristics [of] what data looks like over time supports some of our modeling decisions such as how many years of data to use, and our choice to weight more recent outbreaks more heavily than older ones.

Slide 23: Variance in Outbreak Size

While the previous slide showed that we see different patterns over time, we also see variance in outbreaks size that is worth describing. The top panel in the figure shows the number of illnesses associated with each of the 1,043 outbreaks in our data organized by food categories. So what you can see here is that we see a very wide ranging outbreak size and that these ranges vary across food category. They go from the smallest being 2 to almost 2000 people. These distributions are highly skewed with most outbreaks occurring in the very lower part of the figure, but some outliers that are very large.

One of our concerns about using outbreak illnesses is a measure for estimating attributions is the potential for these very large outbreaks to overly influenced the estimates. What you can see in the bottom panel are these exact same outbreaks, but presented on a log scale. All that's different here is the Y axis. This shows how log transforming outbreak size gets us to a more normalized distribution. This allows us to mitigate some of the influence of outliers. It also makes it a little bit more amenable to some of the modeling that we wanted to do. For this reason, we log transform outbreak size prior to our statistical modeling.

Slide 24: Epidemiological Factors

Outbreak size does not just vary by food category, however. In our exploratory analysis, we identified two other key epidemiological factors related to outbreak size. There were others, but these two really stood out.

First, we found statistically significant differences for most pathogens in outbreak size for single state versus multi-state exposures. For a number of pathogens, we also found differences in size depending on the location in which food was prepared. Both of these findings make some intuitive sense. It's not surprising that multi-state outbreaks tend to be larger than single state outbreaks, or similarly, that outbreaks associated with public establishments or large gatherings of people or institutional populations tend to be larger than those associated with the food prepared in private homes.

Slide 25: Variation in Outbreak Size Across Epidemiological Factors (*Salmonella*)

This figure shows an example of how outbreak size varies by whether exposures occurred in a single state or multiple states. Each line in this figure represents one of the 638 *Salmonella* outbreaks from the data. Red lines are multi-state outbreaks, and blue lines are single state outbreaks. The right-hand size shows the size of each individual outbreak presented here on log scale, while the left shows the mean of that categorical distinction. What you see is that while many single state outbreaks are larger than many multi-state outbreaks, on the whole, the average size of a multi-state outbreak is much larger than the average size of a single state outbreak.

Slide 26: Variation in Outbreak Size Across Epidemiological Factors

This is the same figure exploded to show all four pathogens and the three epidemiological factors that are in our statistical model. Each line is an outbreak with panels showing individual outbreak sizes on the right, in grouped means on the left. The panels from left to right are for *Salmonella*, *E. coli* O157, *Listeria monocytogenes*, and *Campylobacter*. Going down from the top to the bottom is an examination of the single state versus multi-state, the middle shows the location of food preparation, and the bottom row shows food category. Obviously, we are not going to go through all these, what we want you to take away from this is that we do see differences in the average outbreak size when we look across these factors. This is one of the reasons why these factors were included in our model.

Slide 27: Modelling Outbreak Illnesses

How do we model these things? We developed pathogen specific analysis of variance models – or ANOVA models – to estimate the log-transformed number of illnesses based on the three predictors: food category, food preparation, and single state versus multi-state. We actually developed five models as we ended up deciding to model *Salmonella* Enteritidis separately from all other serotypes due to differences across these factors and also because we had enough data to allow us to do that. Each ANOVA model estimates the number of log-transformed illnesses for each outbreak in the dataset.

These estimates are back-transformed to get us model-estimated illnesses. For example, all single state *Campylobacter* outbreaks implicating chicken and in which the food was prepared in a restaurant are given the same model-estimated number of illnesses. When we calculate our attribution fractions, rather than use the number of reported illnesses we use these model-estimated illnesses.

Slide 28: Down-Weighting Older Outbreaks

In addition to that, I mentioned before that we do some temporal weighting. This is maybe a little bit more information than you need about that, but again for the purpose of transparency, we want to make sure we accurately describe what we are doing. We consider recent outbreaks to be more

relevant for estimating current attribution than older outbreaks. However, using only a few years of data can be problematic. It leads to instability in estimates, particularly for pathogen-food category pairs with regular but relatively infrequent outbreaks.

We decided ultimately to use a down-weighting function which outbreaks older than 5 years are given exponentially lower weight. This allows us to keep [these outbreaks] in the model, but also to make our estimates more weighted towards the most recent outbreaks. For a given outbreak, we multiply the number of illnesses by the factor that corresponds to that year of the outbreak and the point on this line. What you see in the figure at the top are actually four different factors with the one that we use in purple as a solid line in the death signs reflecting some other alternatives, just for some perspective. The one we chose which is highlighted in yellow the bottom shows you that over two thirds of the information that goes into our attribution estimates is from the most recent 5 years. With about 5% from data older than 10 years. And about 28% from the intervening years.

Slide 29: Attribution Percentages & Uncertainty

From here, the calculation of attribution fractions is fairly straightforward. Basically, what we are doing here, as with any percentage, is summing the number of illnesses for a food category and divide it by the sum of illnesses across all food categories. So for each pathogen, we take the [set] of reported outbreaks and use the down-weighted model-estimated number of illnesses instead of the [number of reported illnesses].

Speaking more specifically: for a given pathogen, the attribution percentage a specific food category is calculated as the total number of down-weighted model-estimated illnesses due to that pathogen in that food category divided by the number of down-weighted model-estimated illnesses due to that pathogen in all food categories. I should note at this stage that we do combine the two sets of *Salmonella* model predictions into a single pool of *Salmonella* outbreaks for these calculations. We do not estimate attributions separately for SE and other serotypes.

After all [of this] modeling and adjustments we end up with this example here for *Salmonella* in Seeded vegetables we end up with 6159 total adjusted estimated illnesses due to foodborne *Salmonella*; of these 1022 were associated with Seeded vegetables. Therefore, we estimate that 16.6% of foodborne salmonellosis is attributed to Seeded vegetables. In addition to this, we do estimate 90% credible intervals using bootstrap resampling.

Everything that I have presented thus far is the same, methodologically speaking, as what we did for the 2012 estimates

Slide 30: Excluding *Campylobacter*/Dairy

Where our approach differs significantly for 2013 is for *Campylobacter*. In the estimates we present today, we exclude the dairy category from estimates of attribution percentages for *Campylobacter*. In our prior estimates based on outbreaks through 2012, our point estimate was that about 66% of foodborne campylobacteriosis was due to dairy products. This result was driven by the fact that the majority of *Campylobacter* outbreaks in our data are associated with unpasteurized fluid milk. We recognized that raw milk is a high-risk product that causes a great number of outbreaks and outbreak illnesses. However, raw milk is not widely consumed by the general population. Our goal in this effort is to estimate attribution from outbreaks to non-outbreak illnesses – that is, we want to say something about overall diseases in the population based on outbreaks.

We concluded that in this case the high numbers of raw milk outbreaks likely over-represent the Dairy category generally as a source of overall (that is, non-outbreak) campylobacteriosis. This is line with other studies. Case-control studies and other epidemiological studies of sporadic campylobacteriosis assign lower attribution percentages to dairy than to chicken. For example, a U.S. FoodNet case-control study published in 2004 attributed only 1.5% of sporadic campylobacteriosis cases to consumption of unpasteurized milk, compared with 24% to consumption of chicken prepared in a restaurant. Similarly, structured expert elicitation studies estimate the fraction attributable to dairy products to be something less than 10% (and again this is principally raw milk), compared with 30-70% or even over 70% to chicken. We simply do not trust our estimates for *Campylobacter* in Dairy products [when] basing them on outbreak data.

We therefore estimate attribution percentages for all food categories except for Dairy; that is, for *Campylobacter*, these non-Dairy categories sum to 100%. This does not mean that we think the attribution percentage of Dairy products is 0%, only that we do not think outbreak data [should] be used to estimate it. We feel that outbreaks in other categories are more reliable for estimating attribution percentages, which is why we feel comfortable doing this. Removing Dairy outbreaks does result in estimates of *Campylobacter* illnesses and attribution percentages that are more consistent with the published literature, as you will see in the next section.

Speaking of which, now is the time for me to hand things over to my colleague Dr. LaTonia Richardson of the CDC who will present some results.

Recording Timestamp = 36:18

Slide 31: Results

<<LATONIA RICHARDSON>>: OK, thank you, Michael and greetings everyone. So, what I'd like to do now is walk you through the results of running the model.

Slide 32: Foodborne Illness Source Attribution Estimates for 2013

This figure depicts the estimated attribution percentages and the 90% credibility intervals for the 17 food categories for each of the four pathogens, and you'll see that the number of outbreaks [in the analysis] for each pathogen is included right below the pathogen names.

So, just regarding overall themes, one of the things we noticed was that, if you look at the first panel and the last panel at *Salmonella* and *Campylobacter*, you'll see that the illnesses were more widely distributed across the different food categories, whereas when you look at *E. coli* and *Listeria*, you see that there are several categories where the illnesses—there are no illnesses attributed to those categories. And instead, most of the illnesses are attributed to two categories. So, like, for *E. coli* we see most illnesses attributed to Beef and Vegetable Row Crops and then for *Listeria* we see most illnesses attributed to Dairy and Fruits.

Another thing we have to point out, though, is that for all of the pathogens, [...] you'll notice that for most of the food categories, the attribution percentages [credibility intervals] overlap, and this means no statistically significant differences between those categories. So, as an example, if you look at *E. coli*, the attribution --or the credibility intervals--overlap for Beef and Vegetable Row Crops, whereas for Beef and Chicken, the credibility intervals don't overlap, and so this indicates that there is a statistically significant difference in the estimated attribution for Beef and Chicken but no statistically significant difference in the attribution percentage for Beef and Vegetable Row Crops. So, we just wanted to point

that out. This just shows you a general, overall look at the pathogens. What we'd like to do now is to take a closer look at each pathogen one at a time.

Slide 33: *Salmonella*

This figure and the figures that follow for the other pathogens, presents the attribution percentages and the 90% credibility intervals, but this time they are arranged in descending order of the [attribution percentages]. We also present the attribution percentages and the cumulative attribution percentages in a tabular format below the graphic.

So, for *Salmonella*, what we see is that the Seeded Vegetables category had the highest attribution percentage with 16.6%, and this was followed by Eggs at 11.5% and then Chicken at 10.4%. So overall what we saw was that over 75% of the *Salmonella* illnesses were attributed to seven food categories. However, it's important to point out that many of those --the credibility intervals overlap for those categories, and so for many of those the differences were not statistically significant.

Slide 34: *E. coli* O157

For *E. coli*, we saw that the Vegetable Row Crops and Beef categories combined accounted for about 80% of the illnesses for *E. coli*. Again, the credibility intervals for these two categories overlap, so there is no statistically significant differences between them. We also see for *E. coli* that there are five categories with no illnesses attributed to them, and so those include Pork, Eggs, Other Seafood, Grains-Beans, and Oils-Sugars.

Slide 35: *Listeria monocytogenes*

For *Listeria*, we see that the majority of illnesses were attributed to two categories. So, that's similar to what we just saw for *E. coli*, but in this case, we see that Fruits and Dairy account for nearly 90% of the attribution for *Listeria*. Again, the credibility intervals for Fruits and Dairy overlap so there is no statistically significant differences between them. We also see that the credibility intervals are a bit wider in comparison to the credibility intervals for the other food categories, which indicates more uncertainty for the attribution percentages for these categories. We also see for *Listeria* that [there are nine categories] with no illnesses attributed to them, and you see those categories listed here.

Slide 36: *Campylobacter*

For *Campylobacter*, as Michael mentioned, the attribution percentage for Dairy is not presented. So, the attribution percentages that you see here actually reflect non-Dairy illnesses. So, we found that nearly 80% of the non-Dairy illnesses for *Campylobacter* were attributed to five food categories. And then Chicken had the highest attribution percentage with 29.2%. Again, most of the credibility intervals for the attribution percentages overlapped and this indicated no statistically significant differences between them. We also saw that there were no illnesses attributed to Eggs, Grains-Beans, and Sprouts. So, that just gives you an idea of the pathogen-specific results of running the model. What I'd like to do now [is to recap] some of our key findings and discuss some concluding points.

Slide 37: Discussion

[Transition Slide]

Slide 38: Key Findings

So, as we saw on the results, the illnesses for *Salmonella* and *Campylobacter* were broadly attributed across multiple food categories. So, this seems to suggest that food safety interventions designed to reduce illnesses from these pathogens should target a variety of food categories. On the other hand, *E.*

coli and *Listeria* illnesses were attributed to fewer food categories. So, perhaps this suggests more targeted interventions. For *E. coli*, we might expect more targeted interventions in the areas of Beef and Vegetable Row Crops. At the same time, though, we should be mindful of the potential for new and emerging food sources of outbreaks and illnesses and the occurrence of outbreaks due to novel pathogen-food category pairs. An example of this is that earlier this year, we saw an *E. coli* O157 multi-state outbreak due to soy nut butter. In this case, it actually wasn't the Vegetable Row Crops or the Beef, but the category that we would consider Other Produce that was actually the source of that outbreak.

Slide 39: *Campylobacter* Results

For *Campylobacter* results, again, we excluded Dairy from the attribution percentage calculation, and this was for several reasons. As Michael mentioned, most of the *Campylobacter* outbreaks in the database were associated with unpasteurized milk, and that's not widely consumed in the general population. Also, the *Campylobacter*-Dairy outbreaks accounted for 68% of the total *Campylobacter* attribution. So, this seemed to over-represent Dairy as a source of *Campylobacter* illnesses, and this was not consistent with the published literature. When we removed the Dairy outbreaks, we saw the Chicken attribution increase from 9% to 29%. So, this was actually more consistent with the published literature.

Slide 40: Evaluating Our Methods

Consistency with the published literature gave us some assurance about our methods, but we also ran several statistical analyses to evaluate our methods. We compared our results based on model-estimated illnesses to results based on reported illnesses, and we determined that the model appreciably smoothed the variation in outbreak size and it reduced the influence of very large outbreaks. We also conducted a number of sensitivity analyses around the model assumptions and data. For example, we compared different approaches for down-weighting the older outbreaks, like considering different weighting factors. We also considered alternative ANOVA model specifications, such as comparing the results of including versus excluding certain predictors from the model. We also ran analyses regarding sensitivity to outliers, and we considered the effects of including versus excluding suspected etiology outbreaks. So, in the end, we found that our estimates were robust to outliers, but there was one exception, and that is the fact that the *Listeria* attribution estimate was heavily influenced by single large cantaloupe outbreak in 2011. And then as it turns out, later Fruit outbreaks seemed to indicate that Fruit is an increasing concern for *Listeria*.

Slide 41: Limitations

While our methods address a number of issues with outbreak-based foodborne illness source attribution, it is still subject to certain limitations. First, outbreaks in general account for only a small proportion of overall foodborne illnesses, and not all outbreaks get reported to FDOSS. So, among the outbreaks that do get reported, many of them do not implicate a single food, and so that makes it difficult to assign the outbreak to a particular food category. Another limitation of our methods is that for the pathogens with a small number of outbreaks, the outbreaks with a very large illness count can have a [substantial influence] on the attribution point estimate. And that's, again, the example I just provided regarding the 2011 cantaloupe outbreak influencing the *Listeria* attribution estimate for Fruits. Furthermore, our analysis only included 36% of reported outbreaks caused by the four priority pathogens. So, this may not be representative of all outbreaks from these pathogens. Finally, nearly 10% of illnesses in our analysis occurred among institutional populations. So, this includes prisons, hospitals, and schools. These populations, they are generally easier to identify and collect complete data

from, they have fewer food options, and they are just generally not representative of the general population.

Slide 42: Conclusions

In conclusion, our approach addresses a number of issues with outbreak-based foodborne illness source attribution, including the fact that it adopts a food categorization scheme that is [aligned with regulatory needs]. It also addresses biases, it adjusts for outbreak size, and it is generally robust to outliers. Other features of our methods include down-weighting the influence of older outbreaks, and calculated uncertainty around the estimates. However, as we previously discussed, our estimates are still subject to limitations, uncertainties, and biases.

Slide 43: Conclusions, Continued

Our estimates should not be interpreted as suggesting that all foods within a given category are equally likely to transmit pathogens. Further analyses would be needed in order to make these types of claims. We also encourage caution when trying to compare estimates across years. So, a percentage decrease in a certain food category may not necessarily translate to a decrease in the number of illnesses attributed to that category, but rather an increase [in another category]. Again, these estimates are based on percentages, so there is really a zero-sum game here where the increase or decrease in one category directly impacts the other categories. Finally, these estimates are for 2013, and they do not include the recently reported outbreaks from 2014 through 2017.

Slide 44: Conclusions, Continued

Overall, our method is robust, and it also helps to facilitate IFSAC's ability to produce regularly updated, harmonized attribution estimates, and this ensures consistency in the use and interpretation of estimates across agencies. Our estimates can also help to inform food safety decision-making with respect to providing pathogen-specific direction for reducing foodborne illness. Furthermore, annual updates to these estimates not only helps to further IFSAC's efforts to inform and engage stakeholders, but it also furthers the ability to assess whether prevention measures are working. IFSAC is currently engaged in other projects designed to enhance existing attribution efforts and to address the current limitations. So, two examples include projects where we are doing further exploration of *Campylobacter* attribution, and also another project where we are considering including foods assigned to more than one food category.

So, with that, I will now turn it back over to Cary Parker for the question-and-answer period.

Recording Timestamp = 51:22

(Note: Due to technical issues, audio was not recorded for Cary Chen Parker or Joanna Zablotsky Kufel during this portion of the webinar)

Slide 45: Question and Answer Session

<<CARY CHEN PARKER>> [audio not recorded]: [This concludes the presentation portion of the webinar. Many thanks to our four presenters. We will now turn over to any questions you may have. Please use the Q&A box to login your questions and we will take your questions as they come in. If you have not entered your questions yet, please do so at this time so that we may be able to review them. If you have already submitted a question, there is no need to re-enter. Thank you for submitting your questions, I will read each question allowed.]

Question & Answer Session

Recording Timestamp = 52:12

Q1: <<CARY CHEN PARKER>> [audio not recorded]: [The first question is, “Could you further explain the use of suspected vs. confirmed etiologies? What confidence level is used to incorporate these outbreaks?”]

A1: <<MICHAEL BATZ>>: Looking at whether etiology was confirmed or suspected, we did look at some of the reasons further as which outbreaks have been confirmed versus suspected and so forth. Those with suspected etiologies tend to happen earlier in the time period, and the majority of them were due to *Salmonella*. The percent [with] suspected [etiology] differs across pathogens and within our set, *Salmonella* was higher. I don't have all the numbers in front of me right now to give details, but for example, I don't think any of the *Listeria monocytogenes* outbreaks had suspected etiology. We suspect that some of these [outbreaks with suspected etiology] were due to data entry errors earlier in the reporting time period.

We do not do anything to adjust confidence in those outbreaks. That is, we don't down-weight by confirmed versus suspected etiology. But we essentially end up down-weighting a lot of them because of our temporal weighting. Because so many of them occur early in the time frame they end up getting down-weighted by our temporal down-weighting function. We did do sensitivity analysis and exploratory analysis to compare these outbreaks across other characteristics. We really did not see much difference overall, especially looking at the food categories that were associated. We did go through a sensitivity analysis of the final results to see how our estimates and our credibility intervals change depending on whether or not we included the suspected [etiology] outbreaks. There were no rank order changes in the food categories [with highest attribution]. There were slight changes here or there, and as I mentioned earlier, because we are using more outbreaks (it is really only 10% more but it is enough in some cases) for us to get narrower credibility intervals.

I hope this [answers] your question, but certainly if you have additional questions or if I didn't quite understand it, you can email us at the IFSAC email address and we can try to answer more completely off-line.

Recording Timestamp = 54:55

Q2: <<CARY CHEN PARKER>> [audio not recorded]: [Thank you, Mike. The next question, “Should there be a measure to weight the number of illnesses due to a pathogen versus other pathogens, i.e. *Listeria monocytogenes* versus *Salmonella*?”]

A2: <<LATONIA RICHARDSON>>: I think that's an interesting suggestion in terms of weighting different pathogens with respect to other pathogens. [...] Right now, the model is designed to consider the pathogens separately, so, sort of pathogen-specific models. I think that's a great suggestion and something we can definitely consider in the future.

Recording Timestamp = 55:55

Q3: <<CARY CHEN PARKER>> [audio not recorded]: [The next question for Mike, “What percentage of the dairy *Campylobacter* was associated with raw milk or raw dairy products?”]

A3: <<MICHAEL BATZ>>: I was scrambling a little bit here to try to see if I could get some precise numbers to give you, [but couldn’t find them in time]. I will say that it is well over 50% that are raw fluid milk. Another chunk of them are cheese and dairy products made from raw fluid milk, and there was a chunk of unknown, where we don't have enough information to know whether it was pasteurized milk products or not. But I think if you add it all up, it ends up being, of all the foodborne *Campylobacter* outbreaks in the dataset, over two thirds I think end up being due to unpasteurized milk and unpasteurized [dairy] products. I don’t have the numbers in front of me so don’t hold me to them but it’s in that [ballpark], and I would say that of Dairy [category] overall [within *Campylobacter*], I think it ends up being something like 90% due to unpasteurized products. Again, please don't hold me to that! I wasn’t able to pull the numbers up in front of me just now. It's something along those lines. Hope that helps.

Recording Timestamp = 57:20

Q4: <<CARY CHEN PARKER>> [audio not recorded]: [The next question, “Are you going to analyze 2014 through 2017?”]

A4: <<LATONIA RICHARDSON>>: Sure, IFSAC plans to -- our goal is to produce attribution estimates on an annual basis, that's a goal of ours, and we will definitely be sure to keep the public informed about our plans and our progress on meeting this goal.

Recording Timestamp = 58:12

Q5: <<CARY CHEN PARKER>> [audio not recorded]: [We have another question: “Under what category do IQF vegetables fall?”]

<<LATONIA RICHARDSON>>: Sorry, I'm not sure I understand—the IQF Vegetables?

<<CARY CHEN PARKER>> [audio not recorded]: [I think the question was regarding frozen vegetables. Which category does that fall under?]

A5: <<LATONIA RICHARDSON>>: Oh, that’s a good question. We do have categories at the processing levels that define frozen or canned vegetables [or fruit]. So, it depends on sort of the parent category of the vegetable. If you're dealing with frozen melons for instance, then the category would be assigned under frozen, but within the subcategory of melon. I hope that answers the question. So, frozen is a processing category at the lowest level of our categorization scheme.

Recording Timestamp = 59:28

Q6: <<CARY CHEN PARKER>> [audio not recorded]: [We have another question, this will be for Mike. “Why are non-O157 STEC not included in the analysis?”]

A6: <<MICHAEL BATZ>>: The short answer is simply because we don't have enough outbreaks in the data to be able to do that. The numbers are lower than *Listeria*, and it ends up being very difficult -- we certainly couldn't model them statistically using this method. Also, the overall goals of IFSAC are to focus on these four priority pathogens, although we do talk about expanding our focus to other pathogens on a somewhat regular basis.

Recording Timestamp = 1:00:14

Q7: <<CARY CHEN PARKER>> [audio not recorded]: [Next question, LaTonia: “Can you break down the dairy into specific categories (e.g. cheese, ice cream, fluid milk, etc.) regarding *Listeria monocytogenes* outbreaks?”]

A7: <<LATONIA RICHARDSON>>: Yes, we do have subcategories under Dairy for Solid/Semi-solid Dairy Products and then another subcategory for Fluid Milk. So yes, we do have these categories available in our hierarchy, and when the information is available in the outbreak report, we can assign outbreaks to these categories.

Recording Timestamp = 1:00:56

Q8: <<CARY CHEN PARKER>> [audio not recorded]: [Next question: “I work for a large national chain. How would you recommend that we use this data to protect our customers?”]

A8: <<JOANNA ZABLOTSKY KUFEL>> [audio not recorded]: [That's a good question. Obviously, the work that we are doing in IFSAC is meant to be used, and that's why we are holding this webinar and putting these reports out. I would say that it's important to recognize the data points, but it should not be used to the exclusion of other data. But, I think it can help shape [decision-making] where some of the estimates are higher, and just start looking more carefully into those categories of food, and along with other data. Amongst other things we think that it is a good way to start focusing attention on particular areas of interest for foodborne illnesses.]

Recording Timestamp = 1:02:33

Q9: <<CARY CHEN PARKER>> [audio not recorded]: [And in the interest of time, the presenters will take the final question now. The final question says, “If these estimates can inform food safety-decision making and provide direction for reducing foodborne illness, why would you remove raw milk from the report? It seems that should be included if these are the conclusions of the study.”]

A9: <<MICHAEL BATZ>>: I tried to explain this a bit, but we obviously recognize that raw milk is a high-risk product and that there are number of outbreak and outbreak illnesses associated with it. I think that all of us are appropriately concerned about that. The goal of this work though is a very focused one, which is to answer the question of what percentage of overall foodborne diseases due to a pathogen are due to a specific source. We simply find it not realistic to estimate that two thirds of [overall] *Campylobacter* illnesses due to food are due to raw milk. We just don't find that it comports with all of the other evidence that we have. I think we are working to figure out how we can combine [information] to make *Campylobacter* attribution estimates that better reflect the role of raw milk. We really don't feel like we have a good technical approach to solving the problem right now. It is either in or out. As you can see in our prior estimates, we included them for the sake of saying this is what the data says, but it can also be highly misleading to people to suggest that two thirds of [foodborne] campylobacteriosis is coming through raw milk in the whole population of the United States. I really respect this question, and it is one that we wrestle with a lot. I hope moving forward we can come up with some methodological approaches to better deal with this divide, and obviously, continue to work on our messaging around the risks associated with unpasteurized milk products.

Recording Timestamp = 1:05:05

Slide 46: Please contact us

<<CARY CHEN PARKER>> [audio not recorded]: [Thank you. We appreciate your participation in today's webinar and we hope that you will join us again in the future. If we have not answered your questions or if you have any additional questions, please send us an email at IFSAC@FDA.HHS.gov. Please also visit our webpage (<https://www.cdc.gov/foodsafety/ifsac/index.html>) for additional information about IFSAC and our many other projects. We apologize for any audio issues, but you will be able to read the transcript in its entirety when it is posted online. Again, thank you for your attendance, this concludes today's webinar session.]

[Webinar concluded]