Comparing Sporadic and Outbreak-associated Foodborne Illness

[Announcer] This program is presented by the Centers for Disease Control and Prevention.

I'm talking today with Dr. Eric Ebel, a co-author of an article about sporadic and outbreak cases of foodborne illness. Dr. Ebel is a veterinarian and risk analyst with USDA's Food Safety and Inspection Service.

Welcome, Dr. Ebel.

[Eric Ebel] Thanks Sarah. Good to be with you.

[Sarah Gregory] Dr. Ebel, why was this study done?

[Eric Ebel] This project was inspired by a manuscript written by John Painter and others at the CDC – that was eventually published in EID in 2013. That paper estimated the fraction of different microbial illnesses that were attributed to different food categories. For example, it provided an estimate of what fraction of *Salmonella* illnesses were associated with consumption of poultry, beef, pork, dairy, and other general types of foods.

The Painter paper used information from foodborne disease outbreaks to determine their attribution fractions. Outbreaks are suitable for this type of analysis because outbreaks – by definition – include multiple human cases and allow investigators to question those cases to determine which foods they might have in common. Through a process of questioning the people who got sick – and others who might have eaten with them but did not become sick – some outbreaks reveal a most probable food source for the illnesses.

CDC's Foodborne Disease Outbreak Surveillance System includes outbreaks reported by the 50 U.S. states and territories. The system has accumulated a number of outbreaks for most of the foodborne pathogens across a number of years. Those outbreaks provide clues to the foods that seem to be associated with the outbreaks. For example, the fraction of *Salmonella* outbreaks attributed to one of the general food categories is a very appropriate estimate of the contribution of that food to outbreaks. But the question of whether the outbreak estimate is appropriate for the vast majority of illnesses that are not outbreak cases was much more difficult to answer.

So that is how we get to the why of our analysis. The Painter paper assumed that the exposure pathways of foodborne outbreak illnesses were representative of those pathways for all

foodborne illnesses, including outbreak-associated and sporadic illnesses. But, this assumption cannot be tested directly because the food sources of sporadic illnesses typically are unknowable.

[Sarah Gregory] Why are they unknowable?

[Eric Ebel] It's because the food that introduced the pathogen to a single person who gets sick could be any of the foods that the person consumed anytime during the three or more days preceding the onset of their illness. Most cases of foodborne illness are these sporadic or isolated cases and we can rarely home in on which of the foods they consumed are the source of their illness.

Without a clear cause for sporadic cases, we would not be able to determine if those cases had similar exposure pathways to outbreak cases, but we could examine some other characteristics of sporadic and outbreak cases that might reveal if they were different. If we found that sporadic and outbreak cases had different characteristics, then the assumption of similar exposure pathways is less plausible. So, that is what we set out to do.

[Sarah Gregory] Please tell me about the people who conducted this analysis.

Joining us in our effort were statisticians, epidemiologists and other public health scientists from the Interagency Food Safety Analytics Collaboration.

This collaboration was formed in 2011 by the CDC, the FDA, and the Food Safety and Inspection Service of the USDA. It was created to improve coordination of federal food safety analytic efforts – specifically in relation to foodborne illness attribution. This project was one of the Collaboration's first efforts, with the analysis being conducted in 2012 and 2013.

[Sarah Gregory] How was the study conducted?

[Eric Ebel] Well, in lieu of a direct comparison of exposure pathways between outbreak and sporadic foodborne illnesses, we compared selected demographic, clinical, temporal, and geographic characteristics of outbreak and sporadic cases. We did this for illnesses caused by *Campylobacter*, *Escherichia coli* O157, or *E. coli* O157, *Listeria*, and *Salmonella* bacteria. These pathogens were considered a high priority because they are spread commonly through food and cause severe illnesses, and because focused efforts to reduce these bacteria in food can be very effective.

We used data from the CDC-led Foodborne Diseases Active Surveillance Network, referred to as FoodNet, for 2004 through 2011. This is a different surveillance system from the Outbreak Surveillance System I mentioned previously. FoodNet is an active surveillance system that includes 7 states, and selected counties in three additional states. Laboratory data are submitted to the CDC by public health personnel from each FoodNet site. Although most of the cases detected in FoodNet are considered sporadic, some of the cases in FoodNet are classified as associated with an outbreak. Therefore, unlike the Foodborne Disease Outbreak Surveillance System, FoodNet contains information on both outbreak and sporadic cases.

We aimed to determine whether differences exist between outbreak and sporadic cases in terms of 6 selected characteristics reported for laboratory-confirmed *Campylobacter, E. coli* O157, *Listeria*, and *Salmonella* infections in FoodNet. The characteristics examined were: 1) the FoodNet site reporting the case; 2) the year in which a case occurred; 3) the season in which a case occurred; 4) the age of the patient; 5) the sex of the patient; and 6) the hospitalization status of the patient. In other words, whether the patient was hospitalized within a week of specimen collection or not.

[Sarah Gregory] Dr. Ebel, would you tell us how you expected these 6 factors to be related to outbreak or sporadic cases?

Well, we assumed that variability in the fraction of cases that are outbreak cases across the FoodNet sites was most likely explained by differences in State or local resources used to detect and investigate outbreaks among those sites. And that this variability did not necessarily suggest underlying differences in the food sources of sporadic and outbreak cases.

In our analysis, the probability of a case being outbreak-associated did vary significantly across the FoodNet sites for all 4 pathogens studied. Previous research has demonstrated that differences in specimen collection and testing – and outbreak surveillance and reporting practices – contribute to differences across states.

We used a 2-step statistical approach to our analysis of each of the 4 pathogens we examined. The idea for both analyses was to determine if any of the factors were significantly associated with being an outbreak case. In particular, we were interested in the age, sex, and hospitalization factors because these were characteristics of the human cases rather than the surveillance system's ability to detect cases. If our analyses found no apparent association between the probability of being an outbreak case and the factors we studied in the data – especially the age,

sex and hospitalization factors – then we couldn't reject a hypothesis that outbreak and sporadic cases are similar.

We thought that an alternative outcome might be that we find a significant association between the probability of being an outbreak case and one or more of these factors. For example, if we found that hospitalization status was associated with being an outbreak case, then it must be that the occurrence of hospitalizations is different for outbreak cases versus sporadic cases. If outbreak cases have a different frequency of hospitalizations than sporadic cases, then it is possible that the two kinds of cases represent different exposure pathways, or differ in some other substantial way, that calls into question the applicability of outbreak attribution evidence to sporadic cases.

In the first step of the analysis, we used a data classification algorithm to gauge the relative importance of the 6 factors in distinguishing between outbreak and sporadic cases. In a stepwise fashion, we progressively removed the least important factors from the analysis until removal of a factor caused misclassification of the data to worsen, at which point we stopped. Factors that were not eliminated were carried on to the next step.

The second step of the analysis was logistic regression modeling, which, in this setting, identifies the most predictive relationship between being an outbreak case and the factors not eliminated in the first step. We used stepwise model building routines to examine all main effects, and interactions among the factors, to best explain the probability of a case being outbreak-associated.

[Sarah Gregory] And what did you find?

[Eric Ebel] Well, to begin with, the data confirmed that most cases in FoodNet are not outbreak-associated. During the study period of 2004 through 2011, less than 1 percent of the nearly 43,000 *Campylobacter* infections reported in FoodNet were outbreak cases. About 20 percent of around 3000 *E. coli* O157 cases were outbreak-associated. And there were approximately a 1000 *Listeria* cases and 50,000 *Salmonella* cases, of which just over 5 percent were outbreak cases.

The data classification algorithm we used in the first step of our analysis determined that sex and hospitalization status were not important for classifying outbreak and sporadic cases for any of the pathogens, so these factors were excluded from the logistic modeling step. The remaining factors for further consideration were the patient age, year, season, and FoodNet site.

The logistic modeling in the second step of our analysis found a significant association between FoodNet sites and the probability of being an outbreak case for *Campylobacter*. For *E. coli* O157 and *Listeria*, the models that best explained the probability of a case being outbreak-associated included FoodNet sites and the years that cases occurred. For *Salmonella*, the best model included all 4 factors, as well as the FoodNet site by year interactions, the year by season interactions, and the FoodNet site by season interactions.

Closer examination of the interaction terms from the best-fitting *Salmonella* model illustrated complex relationships. For example, in the Oregon, California, and Minnesota FoodNet sites, the proportion of outbreak-associated cases changes substantially across years. Moreover, the directions of changes are inconsistent across the sites. For example, the peaks and troughs of Oregon's proportions across years are nearly the opposite of Minnesota's pattern. Likewise, the interactions were noisy between the season and both the surveillance year and the FoodNet site.

The finding of a significant association between age and the probability of being an outbreak case for *Salmonella* was notable. Closer examination of the results found that it was the group consisting of 0–3 year-old patients that had a markedly lower proportion of outbreak-associated cases relative to the other age groups. This underrepresentation of outbreak cases among the youngest aged cases is what drives the significance of age in the *Salmonella* logistic regression model.

Given the significance of age in the *Salmonella* model, we concluded that applying source attribution estimates derived from foodborne outbreak data to the youngest age stratum of *Salmonella* sporadic cases might not be prudent. At this time, we cannot determine whether the lower frequency of outbreak-associated cases among the youngest of *Salmonella* patients reflects some fundamental difference in the exposure pathways between outbreak and non-outbreak cases, a difference in case detection methods, or something else.

Based on our findings, we concluded that the characteristics of outbreak and sporadic cases captured by FoodNet vary for all 4 pathogens examined. Nevertheless, with the exception of season and age of the patient for *Salmonella* cases, the differences between outbreak and sporadic cases pertain to factors that are probably associated with the inherent variability among complex surveillance systems rather than some fundamental difference in their sources of exposure.

[Sarah Gregory] Your article suggests aggregating data. How would this help?

[Eric Ebel] In the study by Painter, source attribution was estimated by aggregating multiple years of outbreak data across all states reporting outbreaks and applying those to national annual burden of illness estimates. Other researchers have similarly aggregated outbreak data for estimating source attribution.

In our study, the probability of a case being outbreak-associated varied significantly with the surveillance year for *E. coli* O157, *Listeria*, and *Salmonella*. In addition, the season of specimen submission was a significant factor in the *Salmonella* model.

One justification for aggregating outbreak evidence across seasons and years is the need to capture more information than is available from a single season or year. The significant associations we found between the probability of being an outbreak case and the season, year, and state factors suggests that aggregation of outbreak data across time and space might be appropriate to avoid biases introduced by significant local effects. For example, outbreak and sporadic cases might be dissimilar across periods of 1 year but more similar when multiple years are compared.

[Sarah Gregory] How does the information gained from this analysis influence future analyses?

[Eric Ebel] The information gained from this work has been used by the Interagency Food Safety Analytics Collaboration to support using outbreak data to develop harmonized attribution fractions for *Salmonella*, *E. coli* O157, *Listeria*, and *Campylobacter*. Regarding aggregation, our findings crudely suggest these data should probably be aggregated, but not exactly how or what data to aggregate to get stable estimates. We also think our work will inform future attribution estimates for *Salmonella* as analysts think about what to do with younger cases of human salmonellosis. The Collaboration also plans to use this information in future analyses.

Our study was limited to cases that were laboratory-confirmed and reported to FoodNet. Consequently, our conclusions are based on the assumption that people with foodborne illness who did not seek healthcare or did not have a specimen submitted for laboratory testing (and therefore are not reported in FoodNet) are similar to those whose cases were included in our study. Nonetheless, source attribution methods will continue to evolve and will probably include data from multiple study populations. In the future, the type of analysis reported here could be used to examine more detailed case characteristics for similarities and differences between outbreak and sporadic cases. Currently, these types of data are not captured routinely in the U.S.

surveillance systems, but these systems are not static and may provide richer information in the future.

[Sarah Gregory] Thank you Dr. Ebel. I've been talking with Dr. Eric Ebel about his article, Comparing Characteristics of Sporadic and Outbreak-Associated Foodborne Illnesses, United States, 2004–2011, online in the July 2016 issue of Emerging Infectious Diseases at cdc.gov/eid.

I'm Sarah Gregory for Emerging Infectious Diseases.

[Announcer] For the most accurate health information, visit www.cdc.gov or call 1-800-CDC-INFO