

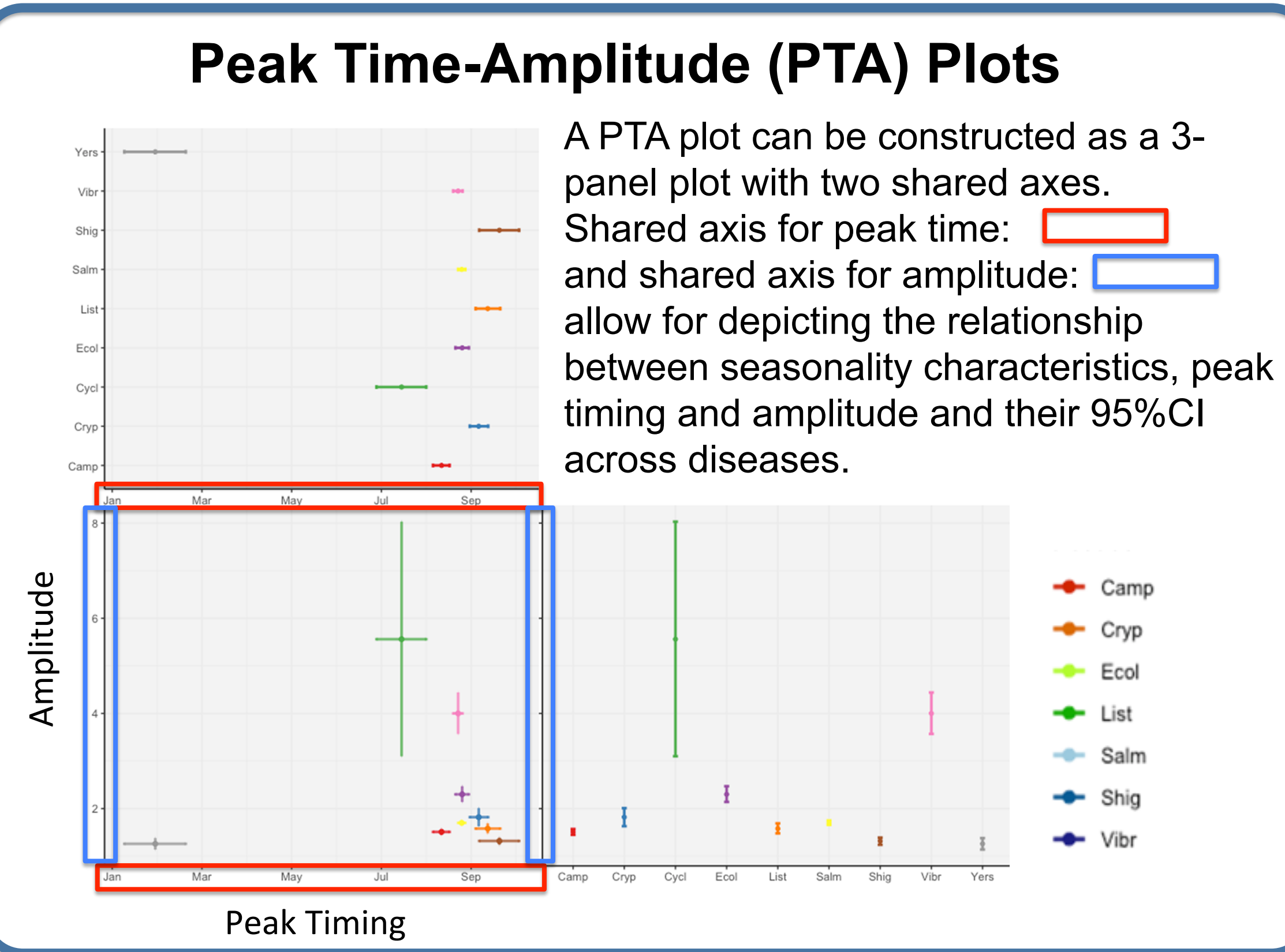
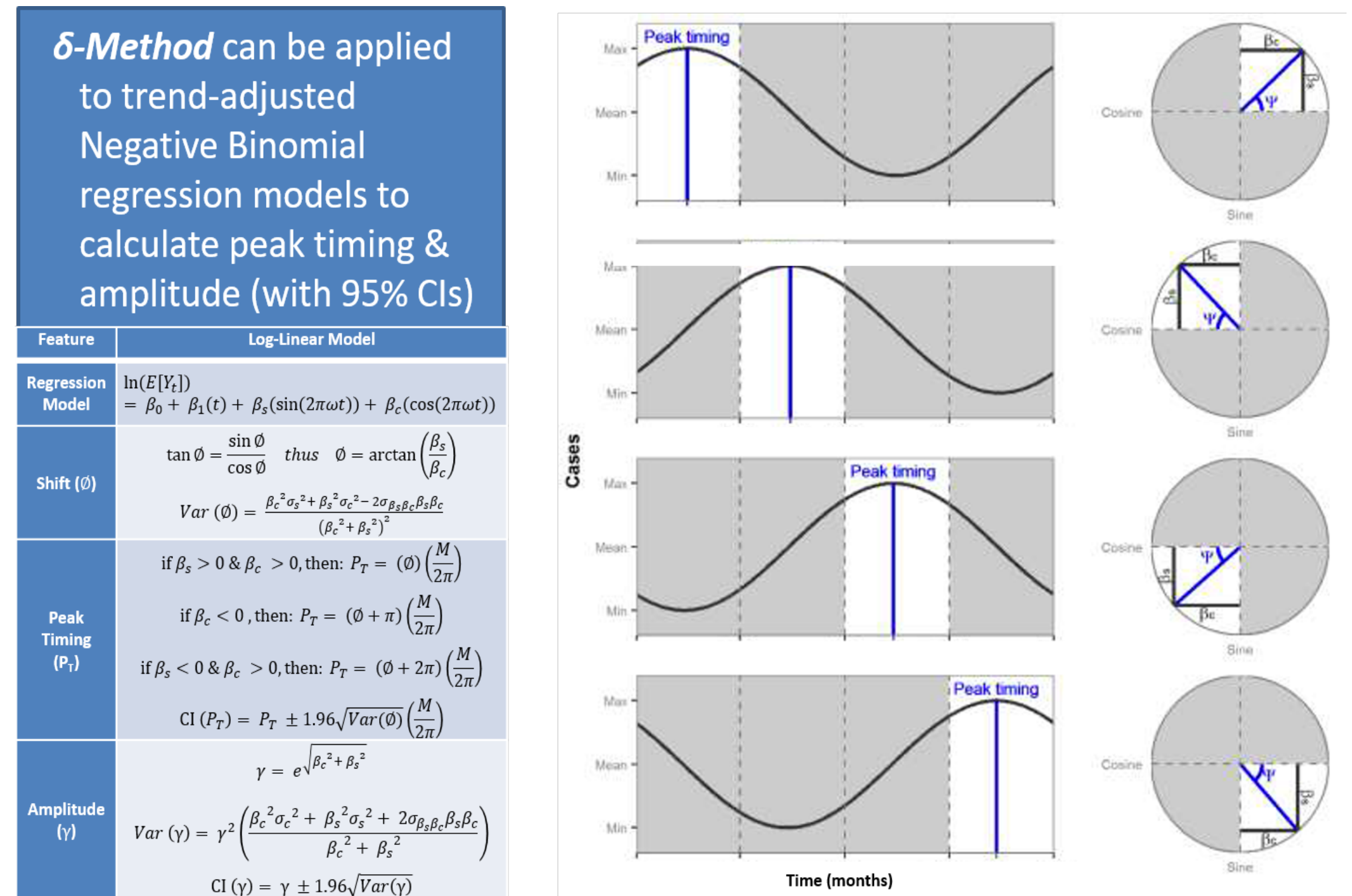
## Objectives

The Centers for Disease Control and Prevention (CDC) estimate that 31 known foodborne pathogens cause 9.4 million cases of these illnesses annually in US. Over 90% of these illnesses are associated with exposure to *Campylobacter*, *Cryptosporidium*, *Cyclospora*, *Listeria*, *Salmonella*, *Shigella*, *Shiga-Toxin Producing E.Coli* (STEC), *Vibrio*, and *Yersinia*. Contaminated products contain parasites typically causing an intestinal illness manifested by diarrhea, stomach cramping, nausea, weight loss, fatigue and may result in deaths in fragile populations. Since 1998, the National Outbreak Reporting System (NORS) has allowed for routine collection of suspected and laboratory-confirmed cases of food poisoning [1]. While retrospective analyses have revealed common pathogen-specific seasonal patterns [2], little is known concerning the stability of those patterns over time and whether they can be used for preventative forecasting. The objective of this study is to construct a calendar of foodborne outbreaks of nine infections based on the peak timing of outbreak incidence in the US from 1996 to 2017.

**Data:** Reported cases were abstracted from FoodNet for *Salmonella* (135115), *Campylobacter* (121099), *Shigella* (48520), *Cryptosporidium* (21701), *STEC* (18022), *Yersinia* (3602), *Vibrio* (3000), *Listeria* (2543), and *Cyclospora* (758). Monthly counts were compiled for each agent and each of 10 states and normalized for captured population.

**Seasonality Characteristics:** Negative Binomial harmonic regression models with the  $\delta$ -method were applied to derive the estimates for peak timing and amplitude and their confidence intervals for each year and for overall study period [2,3] (see Figure 1).

**Figure 1.** An overview of the  $\delta$ -method to derive peak timing and amplitude estimates.



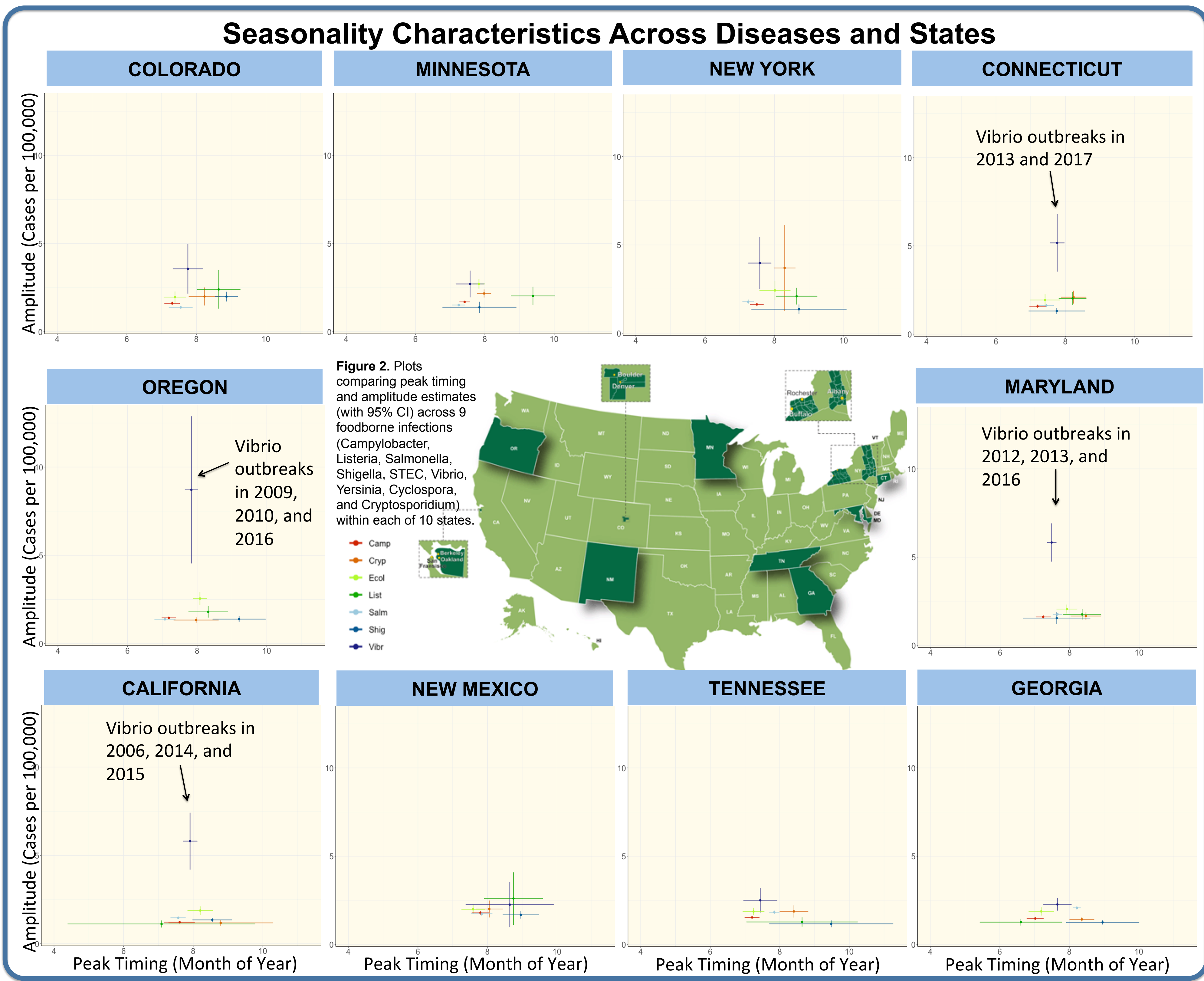
## Acknowledgements

This study was enabled by the National Institutes of Allergy and Infectious Diseases(U19-AI062627 and NO1-A150032). This research is based upon work supported in part by the Office of the Director of National Intelligence (ODNI), Intelligence Advanced Research Projects Activity (IARPA), via 2017-17072100002. The views and conclusions contained herein are those of the authors and should not be interpreted as necessarily representing the official policies, either expressed or implied, of ODNI, IARPA, or the U.S. Government. The U.S. Government is authorized to reproduce and distribute reprints for governmental purposes notwithstanding any copyright annotation therein.

# Foodborne Outbreak Calendar: Applications of Time Series

## An Example for Foodborne infections in the United States (1996-2017)

Ryan Simpson, Margaret A. Waskow, Aishwarya Venkat, Bingjie Zhou, Elena N. Naumova



**Figure 2.** Plots comparing peak timing and amplitude estimates (with 95% CI) across 9 foodborne infections (*Campylobacter*, *Listeria*, *Salmonella*, *Shigella*, *STEC*, *Vibrio*, *Yersinia*, *Cyclospora*, and *Cryptosporidium*) within each of 10 states.

**Overall Trend:** Five infections continue to lead as major causes of outbreaks, exhibiting steady upward trends with annual increases in cases ranging from 2.71% (95%CI: [2.38, 3.05]) in *Campylobacter*, 4.78% (95%CI: [4.14, 5.41]) in *Salmonella*, 7.09% (95%CI: [6.38, 7.82]) in *E.Coli*, 7.71% (95%CI: [6.94, 8.49]) in *Cryptosporidium*, and 8.67% (95%CI: [7.55, 9.80]) in *Vibrio*.

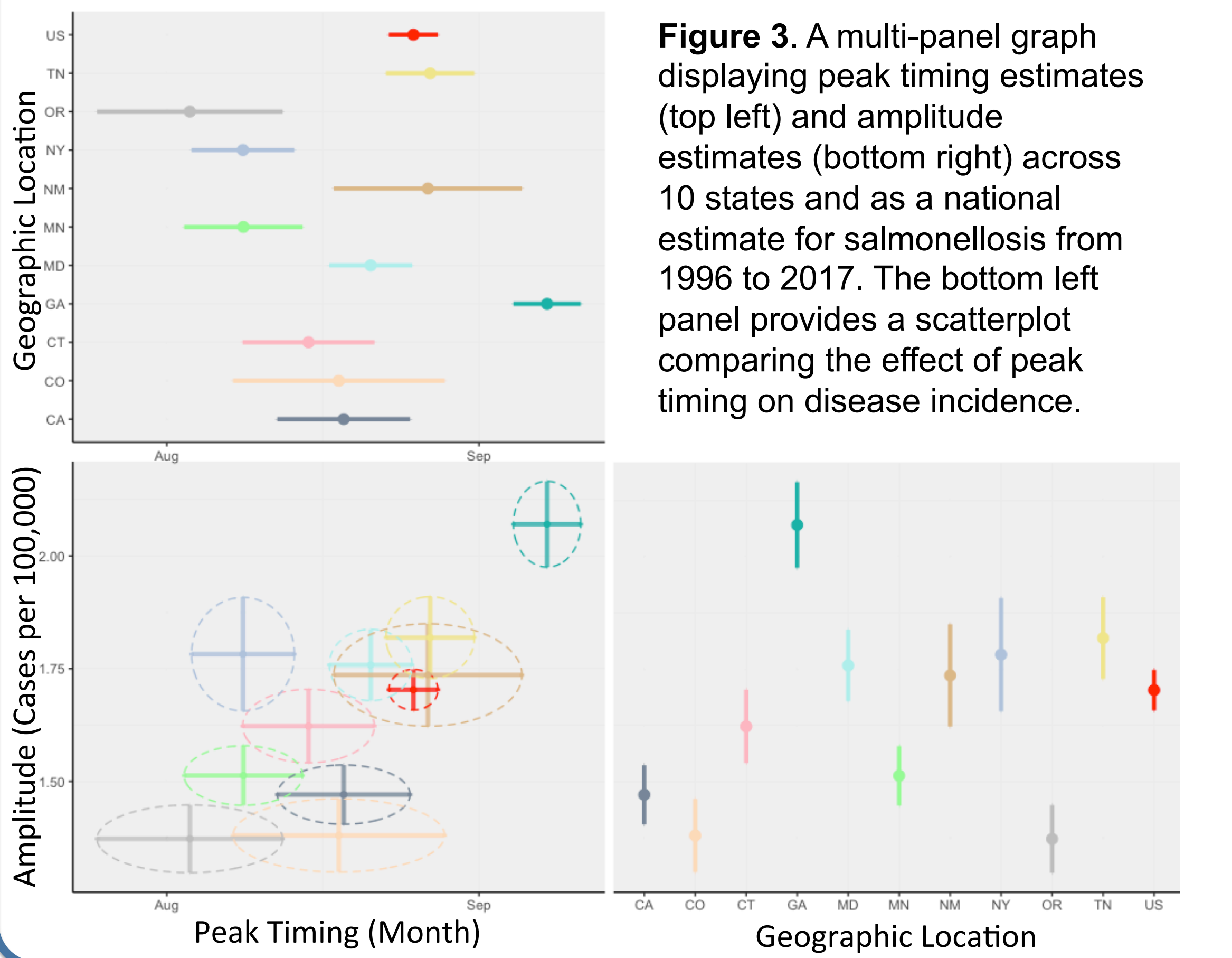
**Seasonality:** Strong synchronization of summer outbreaks were observed, caused by *Campylobacter*, *Vibrio*, *E.Coli* and *Salmonella*, peaking at  $7.57 \pm 0.33$ ,  $7.84 \pm 0.47$ ,  $7.85 \pm 0.37$ , and  $7.82 \pm 0.14$  calendar months, respectively, with the serial cross-correlation ranging 0.81-0.88 ( $p < 0.001$ ). By using harmonic regressions with the  $\delta$ -method

**Peak Timing:** Over 21 years, *Listeria* and *Cryptosporidium* peaks ( $8.43 \pm 0.77$  and  $8.52 \pm 0.45$  months, respectively) have a tendency to arrive 1-2 weeks earlier, while *Vibrio* peaks ( $7.84 \pm 0.47$ ) delay by 2-3 weeks.

## Creating Disease Calendars

Different types of infection calendar structures can be useful for consideration. More commonly observed and cited calendars [2] evaluate differences between diseases for a geographic location (Figure 2). In this design, attention is given to differences between diseases of the same location to better account for differing seasonal patterns across locations. This analysis is limited to a location-by-location assessment to inform scheduling resource allocations (e.g. medications, stockpiling, or, in the advent of limited staff, surveillance testing) between diseases. In contrast, the same calendar format can be applied to compare the same disease across locations (Figure 3). This can be useful to observe variations of a diseases peak timing and amplitude within a larger administrative unit (e.g. across states within the United States). This calendar is more useful for observing changes in the etiology of a disease and adjusting for seasonal outbreaks.

## Seasonality Characteristics Across Geographic Locations: Additional Dimension to Outbreak Calendars



**Figure 3.** A multi-panel graph displaying peak timing estimates (top left) and amplitude estimates (bottom right) across 10 states and as a national estimate for salmonellosis from 1996 to 2017. The bottom left panel provides a scatterplot comparing the effect of peak timing on disease incidence.

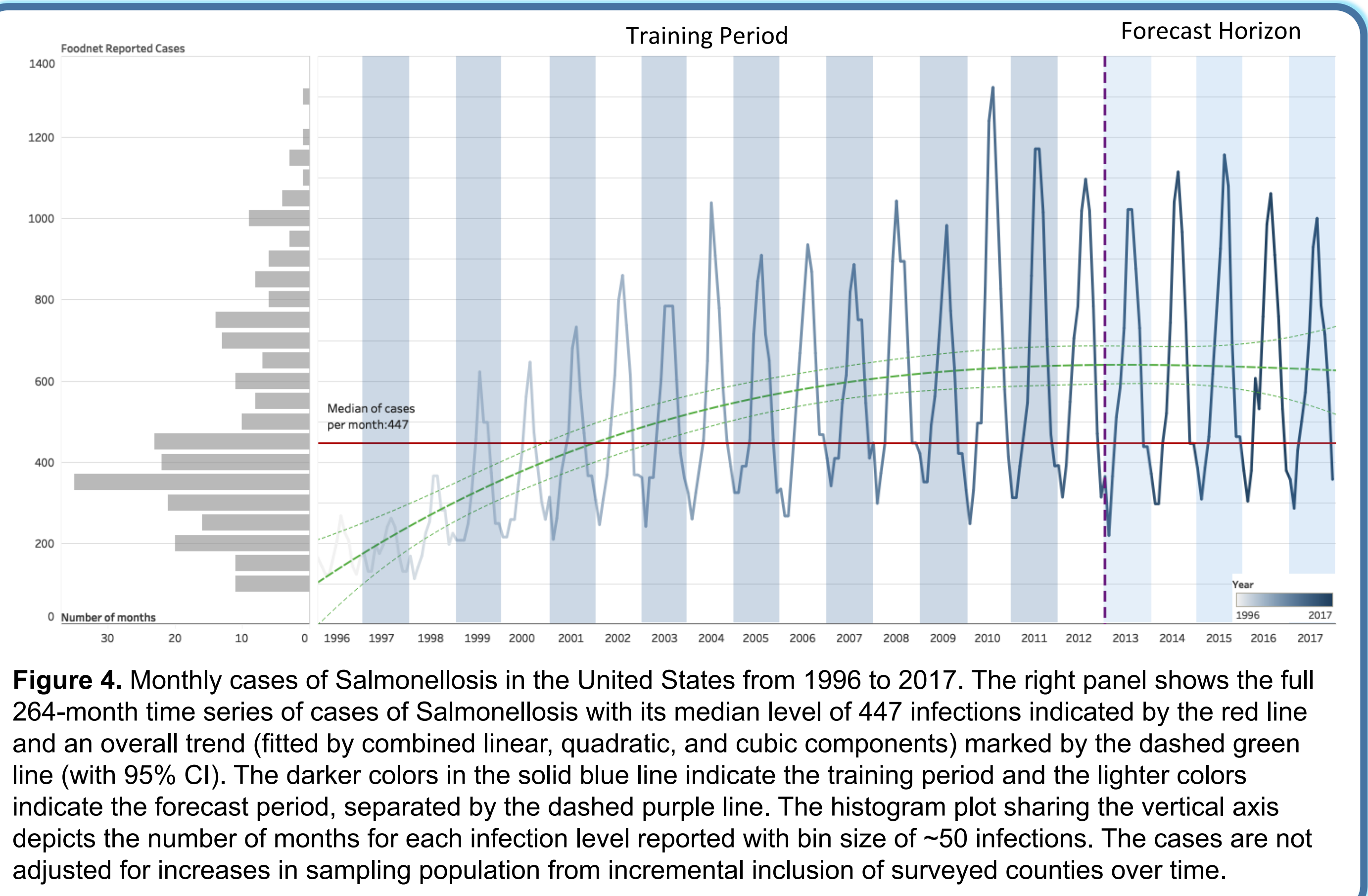
## Forecasting

We used a training period (1996-2012) to build the model to forecast seasonal behavior of salmonellosis (Figure 4). We initially explored the seasonal peak timing and peak amplitude for each year, and their stability. The seasonal peaks of salmonellosis were stable across the training period and for the period intended for prediction (Figure 5); the amplitude appeared to have a slight negative trend. To more adequately adjust for this, we included quadratic and cubic trend values in our forecasted model. \

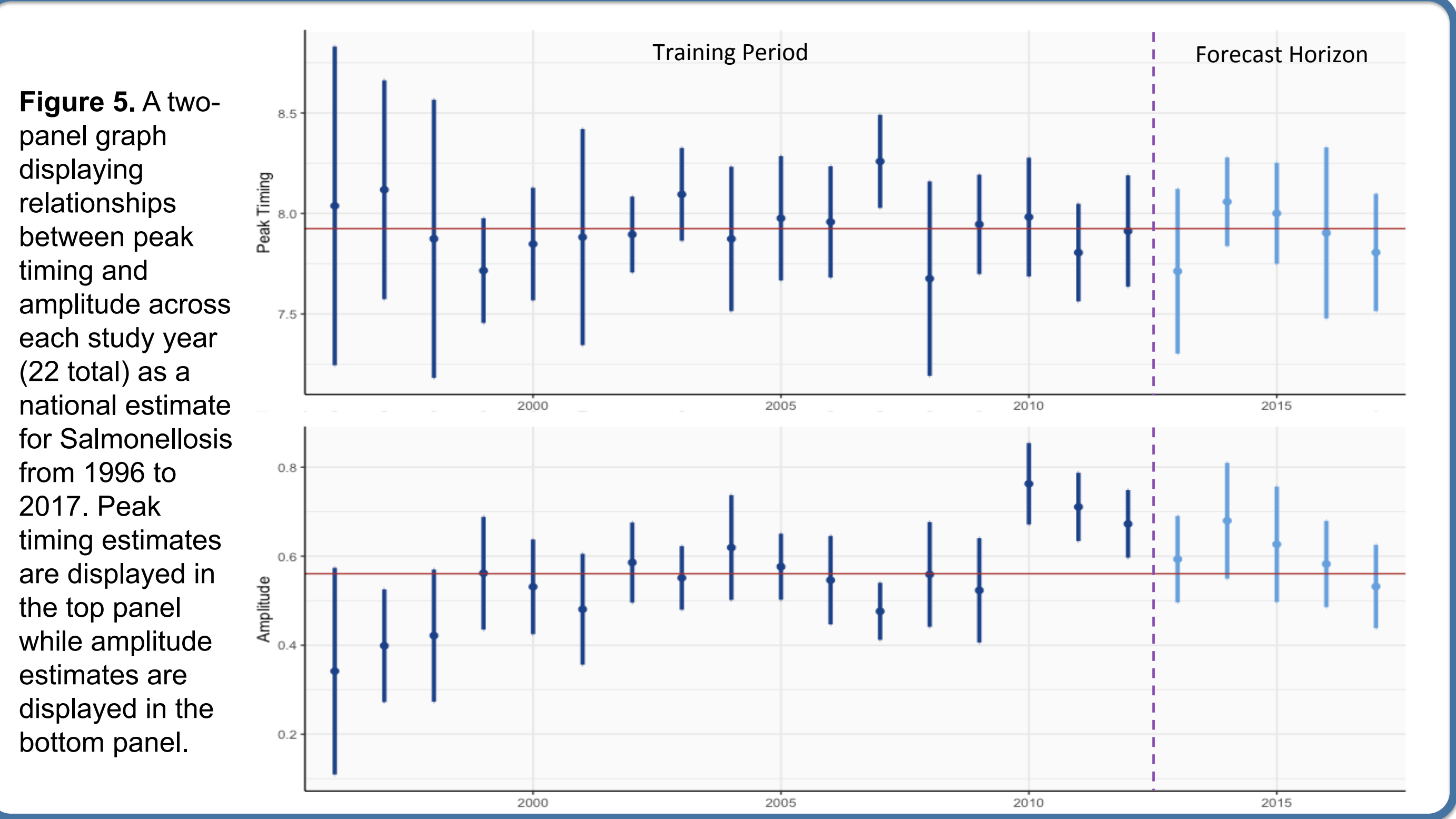
Our forecast (Figure 6) shows an overlay between the observed and forecasted monthly cases from 2013 to 2017. Forecasted values consistently overestimate seasonal nadirs and underestimate seasonal peaks likely due to amplitude trends over the forecast horizon. Peak timing estimates are synchronized across all years. Future efforts will evaluate different model evaluation techniques (e.g. Root Mean Square Error, Akaike's Information Criteria, Bayesian Information Criteria, and percentage differences) to evaluate the accuracy of model predictions.

## References

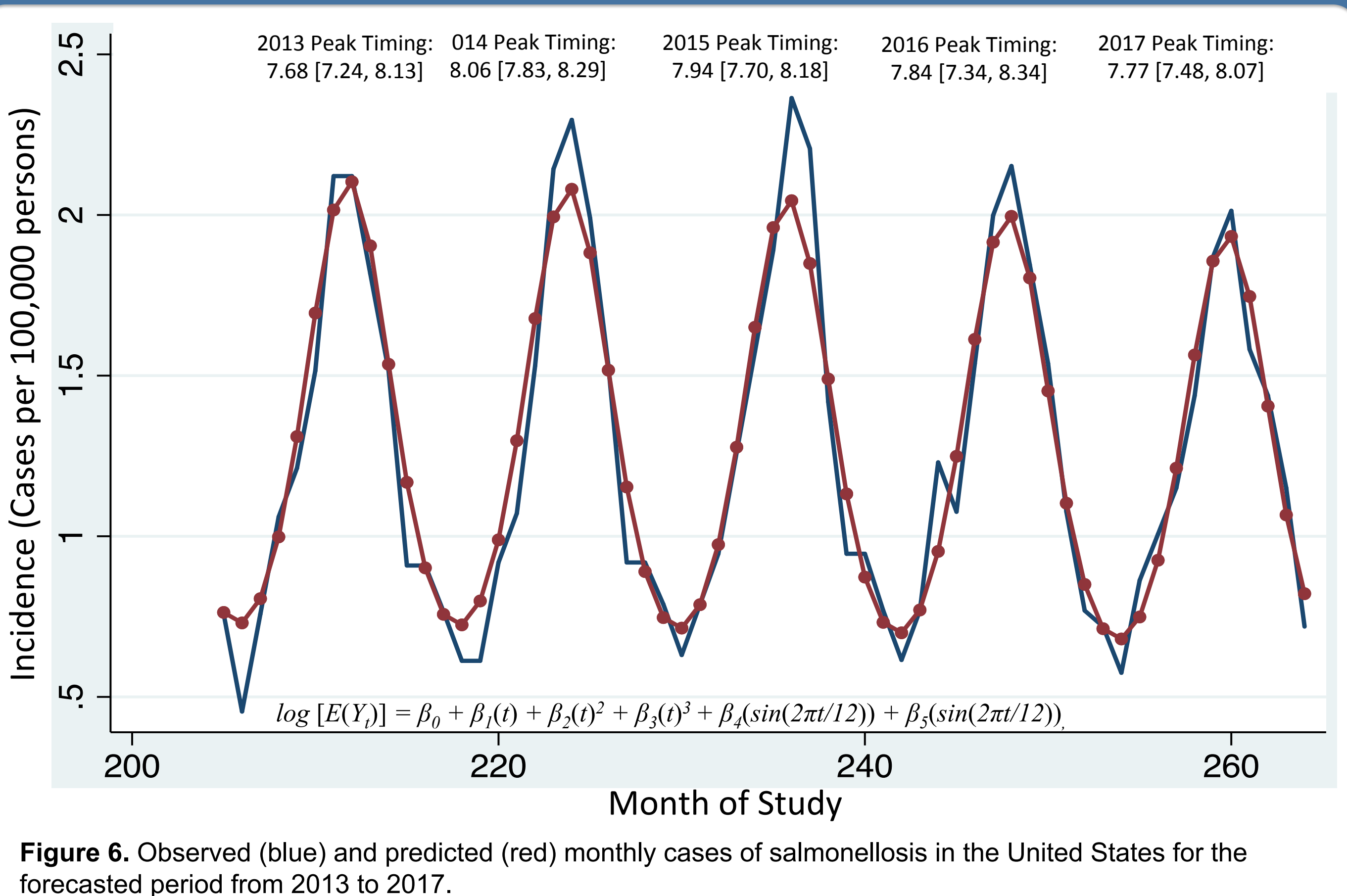
- [1] Centers for Disease Control and Prevention. (2016, Nov). Foodborne Diseases Active Surveillance Network (FoodNet). Retrieved from <https://www.cdc.gov/foodnet/foodnet-fast.html>. Accessed on 31 Dec 2018.
- [2] Martinez, M. E. (2018). The calendar of epidemics: Seasonal cycles of infectious diseases. *PLoS pathogens*, 14(11), e1007327.
- [3] Lofgren, E., Fefferman, N., Doshi, M., & Naumova, E.N. (2007). Assessing seasonal variation in multisource surveillance data: annual harmonic regression. In *NSF Workshop on Intelligence and Security Informatics*. Springer, Berlin, Heidelberg. 114-123.
- [4] Naumova EN & MacNeill IB. (2007). Seasonality assessment for biosurveillance systems. In: Mesbah M, Molenberghs G, Balakrishnan N, eds. *Advances in Statistical Methods for the Health Sciences*. Boston: Birkhäuser. 437-450.



**Figure 4.** Monthly cases of Salmonellosis in the United States from 1996 to 2017. The right panel shows the full 264-month time series of cases of Salmonellosis with its median level of 447 infections indicated by the red line and an overall trend (fitted by combined linear, quadratic, and cubic components) marked by the dashed green line (with 95% CI). The darker colors in the solid blue line indicate the training period and the lighter colors indicate the forecast period, separated by the dashed purple line. The histogram plot sharing the vertical axis depicts the number of months for each infection level reported with bin size of ~50 infections. The cases are not adjusted for increases in sampling population from incremental inclusion of surveyed counties over time.



**Figure 5.** A two-panel graph displaying relationships between peak timing and amplitude across each study year (22 total) as a national estimate for Salmonellosis from 1996 to 2017. Peak timing estimates are displayed in the top panel while amplitude estimates are displayed in the bottom panel.



**Figure 6.** Observed (blue) and predicted (red) monthly cases of salmonellosis in the United States for the forecasted period from 2013 to 2017.

## Conclusion and Future Directions

1. To best construct foodborne outbreak calendars, precise statistical methods are required to identify consistent peak timing of diseases over a prolonged time period as well as inter-annually
2. These calendars must take into consideration variability of disease seasonality across geographic locations for proper application in public health communications and planning
3. Once established, the statistical methods used to develop infection calendars can be utilized for forecasting potential future outbreaks or peak incidence in the following year
4. Greater temporal resolution (daily and weekly compared to monthly) of data resources can provide greater precision for planning purposes
5. These methods can also be applied across databases collecting similar infection information or for a wider range of outcomes (e.g. hospitalizations and deaths) to support planning recommendations.