

1 **An enhanced method for calculating trends in infections caused by pathogens transmitted**
2 **commonly through food**

3 Daniel L. Weller*, Logan C. Ray, Daniel C. Payne, Patricia M. Griffin, Robert M. Hoekstra,
4 Erica Billig Rose, and Beau B. Bruce

5 *Corresponding Author: Daniel L. Weller, 1600 Clifton Rd, Atlanta, GA, qok9@cdc.gov

6 *Enteric Diseases Epidemiology Branch, Division of Foodborne, Waterborne, and Environmental*
7 *Diseases, National Center for Emerging and Zoonotic Infectious Diseases, Centers for Disease*
8 *Control and Prevention*

9 **Abstract**

10 This brief methods paper is being published concomitantly with “Preliminary Incidence and
11 Trends of Infections Caused by Pathogens Transmitted Commonly Through Food— Foodborne
12 Diseases Active Surveillance Network, 10 U.S. Sites, 2016–2021” in *Morbidity and Mortality*
13 *Weekly Reports (MMWR)*. That article describes the application of the new model described
14 here to analyze trends and evaluate progress towards the prevention of infection from enteric
15 pathogens in the United States.

16

17 In 2022, we implemented a new model for estimating trends in illnesses in the ten U.S. sites
18 conducting population-based surveillance for laboratory-confirmed enteric illnesses as part of the
19 Foodborne Diseases Active Surveillance Network (FoodNet). This model was developed to
20 address some limitations of the model used in earlier annual FoodNet reports. Principal
21 improvements include the use of spline regression, an interaction term between site and year, and
22 the treatment of year as continuous (the previous model treated year as a categorical variable).
23 The new method reduces sensitivity to single-year aberrations (e.g., outbreaks), accounts for site-
24 specific trends, and reduces biases toward more populous sites (Figures 1 & 2).

25 The new Bayesian model builds on the previous frequentist model; both are negative binomial
26 models with a log link function and a log offset for the population of each site each year. The
27 previous frequentist model estimated illness count as a function of a categorical year variable and
28 a second categorical variable, called site-entry year, that represented the year a given county in a
29 given site started FoodNet surveillance (*I*). A modified Bayesian version of the previous
30 frequentist model was implemented using an improper flat prior for the fixed effects and a fixed
31 effect of site (instead of site-entry year) to enable comparison between the previous frequentist
32 model and the new Bayesian model described here (Figures 1 & 2).

33 In the new model, illness count is modeled as a function of site and year, which is treated as a
34 continuous variable. The new model allows for the estimation of separate, site-specific trends
35 using penalized thin plate regression splines. These types of splines were selected because they
36 (i) provide computationally efficient, stable, and optimal low-rank approximations compared
37 with thin plate splines, (ii) avoid issues related to knot placement through truncated eigen-
38 decompositions, and (iii) allow model selection using methods dependent on model nesting (2).
39 The priors and the value of the basis function used to represent the smooth term for the thin plate
40 regression splines were set to the package default. For the fixed effects, an improper flat prior
41 was used. For the splines, a half student-t prior was used with the scale parameter dependent on
42 the standard deviation of the transformed response.

43 The new model was implemented with 6 chains and 10,001 iterations using the brms package
44 (version 2.14.0) in R (version 3.6.2). Splines were implemented in the brms model using the `s()`
45 function with the `by` parameter set to site so that separate site-specific splines were fit for each of
46 the 10 sites. The new model generated a posterior predictive distribution of estimated mean log
47 illness counts for each site in each year. We obtained samples from this distribution using the
48 `add_linpred_draws` function in the tidybayes package (version 2.1.1) and exponentiated them.
49 Illness estimates across sites were then summed for each draw-year combination to generate a
50 distribution of illness estimates for the entire FoodNet catchment population for each year during
51 the study period. Median incidence estimates and equal-tailed 95% credible intervals (CrI) were
52 calculated for each year using these catchment-level distributions. The median and 95% CrIs
53 were converted to incidence per 100,000 using the FoodNet catchment population for the
54 specified year. To determine if incidence increased, decreased, or stayed the same, incidence
55 estimates for the most recent year were compared with one or more reference year periods. For
56 example, in the MMWR report published concomitantly with this paper, incidence estimates for
57 2021 were compared with average incidence estimates for 2016–2018. The incidence estimates

58 for 2021 and the reference period were then used to calculate the percentage change for 2021
59 compared with the reference period.

60 **Ethics Statement**

61
62 This activity was reviewed by US Centers for Disease Control and Prevention, National Center
63 for Emerging and Zoonotic Infectious Diseases' human subjects advisor, who determined that
64 this effort was non-human-subjects research and exempt from IRB review.

65 **Data Availability Statement**

66 All data requests should be submitted to the Foodborne Diseases Active Surveillance Network,
67 Enteric Disease Epidemiology Branch, Division of Foodborne, Waterborne and Environmental
68 Disease, National Center for Emerging and Zoonotic Infectious Diseases, US Centers for Disease
69 Control and Prevention.

70

71 **Funding Statement**

72

73 No external funding was used for this project.

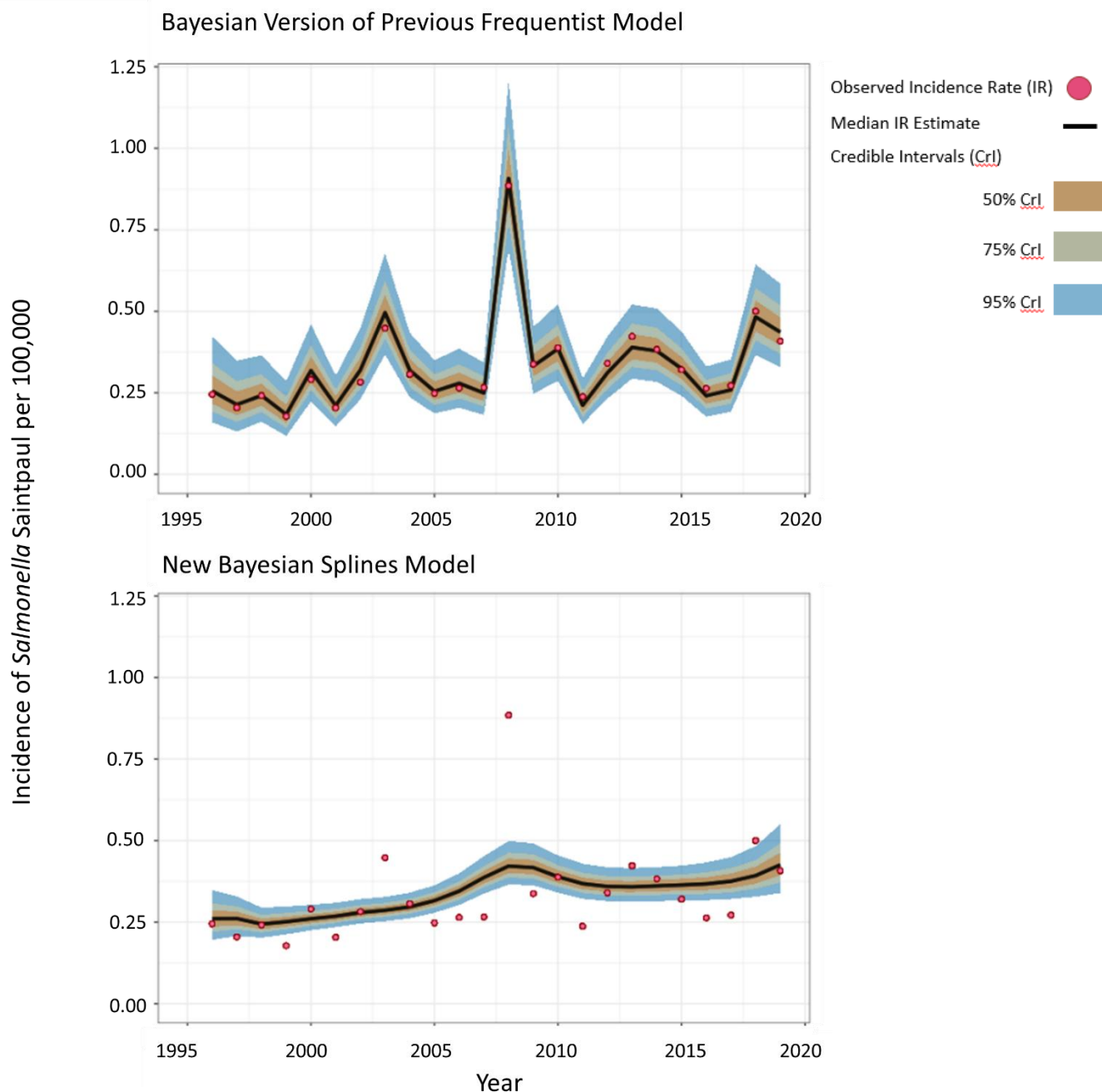
74

75 **References**

76 1. Crim, S. M., Griffin, P. M., Tauxe, R., et al. Preliminary incidence and trends of infection
77 with pathogens transmitted commonly through food – Foodborne Diseases Active
78 Surveillance Network, 10 U.S. sites, 2006-2014. *Morbidity and Mortality Weekly*
79 *Reports*; 2015; 64; 495-499.

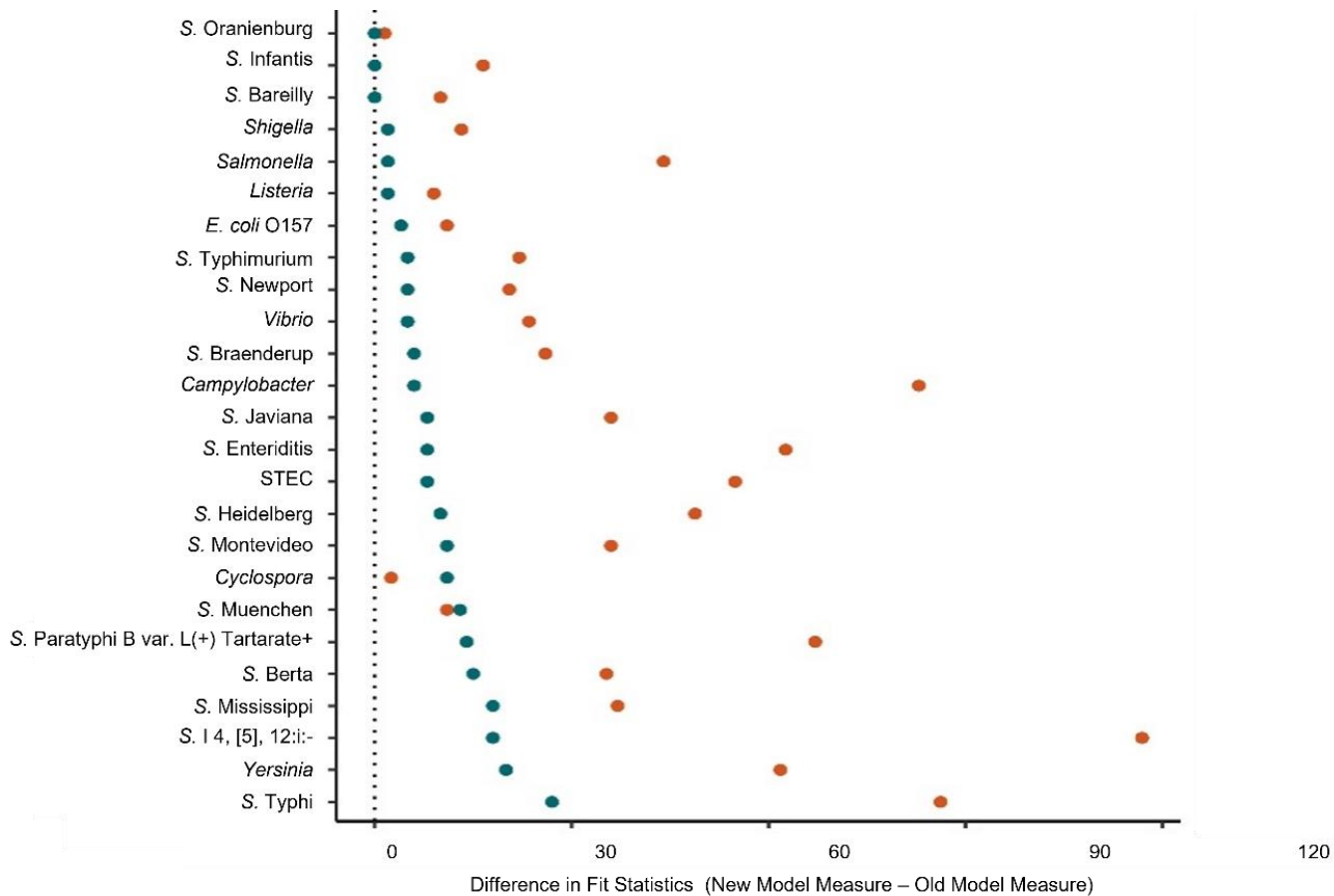
80 2. Wood, S. Thin plate regression splines. *Statistical Methodology*; 2003; 95-114.

81



82
83 Figure 1: Actual (dots) and median estimates (line) of *Salmonella* serotype Saintpaul illnesses
84 per 100,000 population in the FoodNet catchment during 1996 to 2019 according to the previous
85 model^a (top) and the new Bayesian splines model (bottom). The shading represents the 50%,
86 75%, and 95% credibility intervals (CrI). Note how sensitive the previous model is to single-year
87 aberrations, like the 2008 jalapeño and serrano pepper outbreak (vertical blue line), compared
88 with the new model.

89
90 ^aThe model used in prior MMWRs was frequentist; a Bayesian version of that model was created
91 to make outputs that could be compared with those of the new model.



92
 93 Figure 2: Improved fit of the new Bayesian splines model compared with a Bayesian version of the previous frequentist model. Fit
 94 was quantified using two statistical measures: estimated expected log pointwise predictive density (ELPD; orange) and R^{2*} (blue).
 95 The difference was calculated between the new and old models for each measure. Results are shown for all pathogens tracked by
 96 FoodNet, including Shiga toxin-producing *E. coli* (STEC) overall and for *E. coli* O157:H7, specifically, as well
 97 16 *Salmonella* serotypes. Values above 0 (the dotted line) indicate that the new model fit the data better than the previous model;
 98 values below 0 would indicate the opposite. No values fell below 0.
 99
 100 *Differences in R^2 were multiplied by 100.