Banerjee, S.N. (2022). Estimation of high blood lead levels among children in Georgia: An application of Bayesian analysis. *Journal of Environmental Health*, *85*(3), 8–15.

Corresponding Author: Shailendra N. Banerjee, Mathematical Statistician, Division of Environmental Health Science and Practice, National Center for Environmental Health, Centers for Disease Control and Prevention, Atlanta, GA 30341. Email: snb1@cdc.gov.

Note. This supplemental file was submitted by the authors along with the respective peer-reviewed article and has been posted online due to space limitations at https://www.neha.org/jeh/supplemental. The Journal of Environmental Health did not copy edit this file. The authors have provided this supplemental file as an extra resource should the reader want more information.

Detailed Description of the Model Used in the Article, *Estimation of High Blood Lead Levels Among Children in Georgia: An Application of Bayesian Analysis*

Bayesian Model

We used "z" to represent the number of children aged <6 years with BLLs of 5–9 µg/dL in a county in Georgia. Because this is a rare event, we assumed "z" follows a Poisson distribution with parameter "θ" and "m" as the number of children age <6 years who were tested for BLL (equation 1):

$$p(z/\theta) = e^{-(m\,.\,\theta)}(m\,.\,\theta)^z/z! \qquad (1)$$

where θ is the rate of children with BLLs of 5–9 µg/dL

i.e., θ = (children with BLLs of 5–9 µg/dL) / (children tested for BLL)

If θ follows a gamma (α, β) prior

$$\text{i.e., } p(\theta) = e^{-(\beta\theta)}\,\beta^{\alpha}\,\theta^{\alpha-1}/\Gamma(\alpha) \qquad (2)$$

where θ > 0, then posterior distribution of θ is given by

$$p(\theta/z) = p(z/\theta) \text{ X } p(\theta)/p(z)$$

$$\text{i.e., } p(\theta/z) = e^{-(m\,.\,\theta)}(m\,.\,\theta)^z \text{ X } e^{-(\beta\theta)}\,\beta^{\alpha}\,\theta^{\alpha-1}/z!\,\Gamma(\alpha)\,p(z)$$

$$\text{i.e., } p(\theta/z) \propto e^{-\theta(\beta+m)}\,(\theta)^{z+\alpha-1} \text{ X constant} \qquad (3)$$

So, the posterior is also a gamma (α_1, β_1) with

$$\alpha_1 = z + \alpha \text{ and } \beta_1 = \beta + m \qquad (3a)$$

If we assume that the prior information about parameter θ, the rate of children with BLLs of 5-9 μg/dL, can be obtained from a small group of counties in Georgia, who we believe has the same rate (θ) of 5-9 μg/dL BLLs among children aged <6 years, then the posterior for θ can be estimated from equation (3).

We suppose z_j is the number of children aged <6 years with BLLs of 5–9 μg/dL among x_j children from county "j". Then assuming it follows a Poisson distribution, we have the following:

$$p(z_j/\theta) = e^{-(x_j\theta)} (x_j\theta)^{z_j}/z_j! \qquad (4)$$

where θ is the same as defined earlier.

So, the likelihood function for n counties with the same parameter θ is given as follows:

$$L(\sum z_j/\theta) = e^{-(\sum x_j\theta)} \prod (x_j\theta)^{z_j}/z_1!\ z_2!\ .....z_n! \qquad (5)$$

So,

$$L(\sum z_j/\theta) \propto e^{-(\sum x_j\theta)} (\theta)^{\sum z_j} \qquad (6)$$

If for all these n counties we assume that θ follows a non-informative prior $1/\theta$, i.e., $p(\theta) = 1/\theta$, then from equation (6), the posterior distribution of θ is given by the following:

$$p(\theta/\sum z_j) \propto e^{-(\sum x_j\theta)} (\theta)^{\sum z_j} \cdot 1/\theta$$

$$\text{i.e., } p(\theta/\sum z_j) \propto e^{-(\sum x_j\theta)}(\theta)^{\sum z_j - 1} \qquad (7)$$

This is a gamma (α_2, β_2), where

$$\alpha_2 = \sum z_j \text{ and } \beta_2 = \sum x_j \qquad (8)$$

Here, $\sum z_j$ is the shape parameter and $\sum x_j$ is the rate parameter of this gamma distribution, where, z_j is the number of children aged <6 years with BLLs of 5–9 μg/dL in county "j" and x_j is the number of children tested for BLL in county "j". We assume that the rate of children with BLLs of 5–9 μg/dL among children aged <6 years in these counties is similar as that in the targeted county. We then can use known α and β from equation (8) in equations (2) and (3) to evaluate the prior and posterior distributions of the parameter θ in the targeted county.

The joint distribution of data z and the parameter θ are given by the following:

$p(z, \theta) = p(\theta) \times p(z/\theta)$, and also

$p(z, \theta) = p(z) \times p(\theta/z)$

Thus, $p(z) \times p(\theta/z) = p(\theta) \times p(z/\theta)$, giving

$$p(z) = p(\theta) \times p(z/\theta)/p(\theta/z) \qquad (9)$$

Here, $p(\theta)$ and $p(\theta/z)$ are the known prior and posterior distributions, respectively, of the parameter θ. Thus, $p(\theta)$ is a gamma density with the known shape and rate parameters from equation (8). Similarly, $p(\theta/z)$ is a gamma density with known shape and rate parameters from equations (8) and (3a).

Assuming that $p(z/\theta)$ is the sampling distribution of data in the targeted county, we can estimate the predictive density p(z) of z in the targeted county from equation (9), before any data are observed, where $p(z/\theta)$ is a Poisson density with known mean "$m\theta$," as shown in equation (1).

If our model assumptions for sampling distribution of data and prior density are valid, we can check the validity of the observed values of the number of children aged <6 years with BLLs of 5–9 µg/dL in the targeted county.