Journal of Data Science **3**(2005), 47-68

A Comparison of the Posterior Choropleth Maps for Disease Mapping

Balgobin Nandram[1], Jie Liu[2] and Jai Won Choi[1]
[1] *Worcester Polytechnic Institute and* [2] *Research Triangle Institute*

Abstract: In Bayesian analysis of mortality rates it is standard practice to present the posterior mean rates in a choropleth map, a stepped statistical surface identified by colored or shaded areas. A natural objection against the posterior mean map is that it may not be the "best" representation of the mortality rates. One should really present the map that has the highest posterior density over the ensemble of areas in the map (i.e., the coordinates that maximize the joint posterior density of the mortality rates). Thus, the posterior modal map maximizes the joint posterior density of the mortality rates. We apply a Poisson regression model, a Bayesian hierarchical model, that has been used to study mortality data and other rare events when there are occurrences from many areas. The model provides convenient Rao-Blackwellized estimators of the mortality rates. Our method enables us to construct the posterior modal map of mortality data from chronic obstructive pulmonary diseases (COPD) in the continental United States. We show how to fit the Poisson regression model using Markov chain Monte Carlo methods (i.e., the Metropolis-Hastings sampler), and obtain both the posterior modal map and posterior mean map are obtained by an output analysis from the Metropolis-Hastings sampler. The COPD data are used to provide an empirical comparison of these two maps. As expected, we have found important differences between the two maps, and recommended that the posterior modal map should be used.

Key words: COPD, Metropolis-Hastings sampler, mode-Hessian, optimization, Poisson regression.

1. Introduction

In Bayesian analysis of mortality rates it is standard practice to present the posterior mean rates in a choropleth map. A natural objection against the posterior mean map is that it is not the "best" representation of the mortality rates (i.e., the most probable map should be presented). From a Bayesian point of view one should really present the map that has the highest posterior density over the

ensemble of areas in the map (i.e., the coordinates that maximize the joint posterior density of the mortality rates). Thus, unlike the posterior mean map, the posterior modal map maximizes the joint posterior density of the mortality rates. In a Bayesian analysis, a scientist can generate a large number of maps from an output analysis of an appropriate sampling based method. By simply presenting the posterior mean map useful information may be lost. Our objective is to develop a method to construct the map that has the highest posterior density.

The first known choropleth map was constructed on education rates in France in 1826 by Charles Dupin, an education reformist not a cartographer. But the word "choropleth" had to wait over one hundred years before it was invented in 1938 by Wright, a cartographer from the American Geographical Society in New York City. A choropleth is an areal symbol and "choropleth" means "area" and "fullness, quantity" as Wright said. Technically a choropleth is based on a stepped statistical surface identified by colored or shaded areas called chorograms (e.g., statistical or admisistrative areas). One of the most common forms of mapping data today is the choropleth map, in which each area (e.g., state or county) is shaded according to the characteristic (e.g., mortality rate, crime rate, income, rain fall). Areas with higher values of the characteristic are shaded more darkly and vice versa. In the United States of America choropleth maps are used in almost all applications, even in the daily newspapers and on television. Three characteristics of these maps are (a) the value at specific area, (b) overall pattern on the map and (c) pattern on one map as compared with the pattern on other maps. For each area uniform distribution is assumed: the spatial unit used for shaded mapping (same color) is the smallest detail that the shaded map can represent. Within this unit the variable being mapped is uniformly distributed. If the areas are too large, this type of mapping can hide important variation in these areas. Too small units may, however, introduce visual noise. Aggregating these units to larger ones may better reveal a visual pattern of the data. It is important to choose the right classification method and there are two main considerations (a) the interpretation skills of an expected user and (b) the best classification method to represent particular data. Classification of the areas can be done by forming intervals across the range of the data. For example, these intervals can be equal widths, formed from quantiles or using natural breaks. In our work on mortality data we generally use quantiles (e.g., quintiles) and the areas in the higher quantiles get deeper colors or heavier shades in grey scale.

Recently, there has been increased interest in estimating mortality rates for small geographical areas. Models and methods of analysis on rates are abundant (see Nandram, Sedransk and Pickle 1999, Nandram, Sedransk and Pickle 2000, Waller, Carlin, Xia and Gelfand 1997). For the purpose of constructing the posterior modal map, we use a Bayesian hierarchical model recently discussed by

Christiansen and Morris (1997). This is a Poisson regression model that has been used to study mortality data and other rare events when there are occurrences from several areas. The model utilizes a form in which there are convenient Rao-Blackwellized estimators of the mortality rates. See Nandram (2000) for a review of this model. Christiansen and Morris (1997) describe a hierarchical Bayesian model for heterogeneous Poisson counts under the exchangeability assumption, called Poisson regression interactive multilevel modeling (PRIMM). They have made some analytical approximations which are very accurate, and it is important to note that these approximations avoid the use of sampling based methods such as Markov chain Monte Carlo (MCMC) methods. A sampling based method helps us to find the rates that make the posterior density over the entire ensemble the highest. This is a desirable approach in a Bayesian analysis.

It is pertinent to point out a related problem. In many disease mapping problems, presenting the mean rates is a natural and effective practice. However, in the analysis on rare events (e.g., mortality rates of COPD), this often turns out to be misleading because the distributions of such rare events are usually skewed. This can produce a difference between the means and the modes, and presenting means will fail to give us accurate information. We give a simple illustration. Let us denote the mortality rate for an area by R and suppose $R \sim \text{Gamma}(\alpha,\ \beta)$, a natural choice for the distribution of mortality data. Then, the mean of R is $\mu_1 = \alpha/\beta$ and the mode of R is $\mu_2 = (\alpha - 1)/\beta$. Now, suppose $\alpha = 2$ and $\beta = 10^{-4}$, then $\mu_1 = 2 \times 10^{-4}$ and $\mu_2 = 10^{-4}$ (i.e., the mean rate is twice as large as the modal rate, a significant difference). Yet in this example the mean is not so plausible as the mode. But if $\alpha \gg 1$, the difference between the mean and the mode will be negligible. However, we believe that the map based on the mode should really be the one drawn.

For convenience we denote the number of health ervice areas (HSAs) by $\ell = 798$. Let $\underset{\sim}{\lambda}$ denote the ensemble of mortality rate parameters and $\underset{\sim}{d}$ represent the data. That is, $\underset{\sim}{\lambda} = (\lambda_1, \ldots, \lambda_\ell)$, and the data consist of the deaths $\underset{\sim}{d} = (d_1, \ldots, d_\ell)$ and the population sizes $\underset{\sim}{n} = (n_1, \ldots, n_\ell)$ which are known. We ignore the covariates momentarily. In the Bayesian view, given $\underset{\sim}{\lambda}$, the deaths have a distribution; given hyperparameters, $\underset{\sim}{\lambda}$ have a distribution (hyperparameters are parameters of this distribution); and finally the hyperparameters have a distribution. This is a hierarchical Bayesian model. Note that unlike non-Bayesian inference $\underset{\sim}{\lambda}$ is a random vector. Then, using Bayes' theorem and some integration, the joint posterior density of $\underset{\sim}{\lambda}$ is $\pi(\underset{\sim}{\lambda} \mid \underset{\sim}{d})$. Note that the key idea in Bayesian statistics is that all information about $\underset{\sim}{\lambda}$ resides in $\pi(\underset{\sim}{\lambda} \mid \underset{\sim}{d})$. Also, it is important to note that the components of $\underset{\sim}{\lambda}$ are correlated a posteriori. The *posterior mean* map is obtained by drawing the choropleth map for the posterior means of $\lambda_i,\ i = 1, \ldots, \ell$. Clearly, this ignores the inherent correlation among

the components of $\underset{\sim}{\lambda}$, and this is one additional obvious short comings of the posterior mean map. One needs to construct a map simultaneously across the areas (i.e., incorporate the correlation). However, the *posterior modal* map is a map that plots the joint posterior mode over the surface $\pi(\underset{\sim}{\lambda} \mid \underset{\sim}{d})$ providing a point in ℓ dimensional space that includes this correlation (i.e., the synergism or antagonism over the components of $\underset{\sim}{\lambda}$). The Bayesian view states that the point that has the highest posterior density (i.e., the posterior mode) should be used as a representative value; otherwise the entire posterior density should be presented. But, it is clearly impossible to present the entire posterior density in a high dimensional space such as in the case of drawing choropleth maps. When there is skewness, the mean is not a high density point, but the mode is.

Optimization of a function in a high dimensional space is a very complex task, but it is easier if the function is a density function as in our application. Simulation methods (e.g., simulated annealing) are attractive because constraints on both the regularity of the function and on the function itself can be largely ignored. Of course, if these constraints can be included, there will be an improvement in the optimization process. Monte Carlo optimization uses the idea that if the function can be "transformed" (not needed in our application) to a probability density function and a random sample can be obtained from it, then one can maximize the original function. One can simply find the sample member among all sample members where the function is the highest to serve as a surrogate for the maximum; see Chapter 5 of Robert and Casella (1999). In our application we need to maximize a posterior density function in a high dimensional space.

Then, what is our approach to construct a posterior modal map? The answer is simply how to maximize the posterior density $\pi(\underset{\sim}{\lambda} \mid \underset{\sim}{d})$; but this is a complex task because of the high dimensionality. Our method for doing this relies on the output from a MCMC method (i.e., a sampling based method). The posterior density $\pi(\underset{\sim}{\lambda} \mid \underset{\sim}{d})$ does not exist in closed form. This is a marginal density which is obtained by integration over the parameters of the joint posterior density of $\underset{\sim}{\lambda}$ and the parameters. Moreover, when this is done, $\pi(\underset{\sim}{\lambda} \mid \underset{\sim}{d})$ becomes extremely complex (one can only imagine it). It is now a routine calculation to use a MCMC method to obtain a sample from the entire joint posterior density. For many real applications using Bayesian statistics this is the preferred route, and it has led to the solutions of many complex problems that are otherwise intractable. In our procedure we let $\underset{\sim}{\lambda}^{(h)},\ h = 1, \ldots, M$ denote the ensemble of mortality rates for $M \approx 1000$ samples from a sampling based method; so there are M choropleth maps available to us. Which map should be presented? It should be clear by now that from a Bayesian perspective we should present the map which makes the posterior density $\pi(\underset{\sim}{\lambda} \mid \underset{\sim}{d})$ the highest (i.e., joint mode).

The method for constructing the posterior map is illustrated using mortality

data from chronic obstructive pulmonary diseases (COPD) in the continental United States. The dataset was compiled by the National Center for Health Statistics (NCHS), Hyattsville, Maryland. It contains the number of deaths, population sizes, and a set of potential explanatory covariates for the 798 health service areas (HSAs) in the contiguous 48 states. In our study we also tried to link the mortality rate to the potential explanatory variables (covariates). These covariates include smoking history, population density, elevation, annual rainfall level, summer rainfall level, average income level, and college student ratio. Here lung cancer rate within a HSA is used as a surrogate for smoking history. Previous study by Nandram, Sedransk and Pickle (2000) shows that for older white males, wmlung (white male lung cancer rate), sqrtpopd (square root of population density), sqrtelev (square root of elevation) and arain (annual rainfall level) are significant at a 5% significance level. Our regression analysis is consistent with their result, so we include these four covariates in our study.

Data were collected during 1988-1992 with 10 age classes being identified. In our study, we focus on the age classes which contain age 65+ (65 and older). This age group is of particular interest because COPD occurred much more often, and is a frequent cause of death of our retirees. We show how to fit the Poisson regression model using Markov chain Monte Carlo methods (i.e., the Metropolis-Hastings sampler), and obtain both the posterior modal map and posterior mean map by an output analysis from the Metropolis-Hastings sampler. Using the COPD we compare the two types of maps empirically.

Our main objective of this paper is to show how to construct the posterior modal map and to compare it with the posterior mean map of the COPD data for white males age 65+. We describe the Poisson-gamma regression model (Christiansen and Morris 1997) in Section 2, and for illustration we show how well it fits the COPD data using the Metropolis-Hastings algorithm. In Section 3 we describe how to construct the posterior modal map using an output analysis from the Metropolis-Hastings sampler. In Section 4 we present a data analysis on the COPD data and compare the posterior mean map and the posterior modal map. Section 5 has concluding remarks.

2. A Hiearchical Bayesiab Regression Model

In this section we describe the Poisson-gamma hierarchical Bayesian model, how to fit it using the Metropolis-Hastings sampler and the goodness-of-fit using a Bayesian cross validation analysis. Let λ_i denote the mortality rate for the ith HSA, $i = 1, \ldots, \ell$ where $\ell = 798$.

The observations consist of the number of deaths d_i and the population size n_i of the ith HSA, $i = 1, \ldots, \ell$. To link the d_i and the n_i to the mortality rates

λ_i, we assume that

$$d_i \mid \lambda_i \overset{ind}{\sim} \text{Poisson}(n_i\lambda_i), \quad i = 1, \ldots, \ell. \tag{2.1}$$

Under this model the maximum likelihood estimator of λ_i is $r_i = d_i/n_i$, $i = 1, \ldots, \ell$, the observed mortality rate.

It is standard to estimate the λ_i by "borrowing strength" across the 798 HSAs. Thus, letting $\underset{\sim}{x}_i = (1, x_{i1}, \ldots, x_{i,p-1})'$ denote the vector of $(p-1)$ covariates and an intercept, we assume that

$$\lambda_i \mid \alpha, \underset{\sim}{\beta} \overset{ind}{\sim} \text{Gamma}(\alpha, \alpha e^{-\underset{\sim}{x}_i'\underset{\sim}{\beta}}). \tag{2.2}$$

Observe that in this model $\log(E(\lambda_i \mid \alpha, \underset{\sim}{\beta})) = \underset{\sim}{x}_i'\underset{\sim}{\beta}$ and $\sqrt{\alpha}$ is the coefficient of variation of the λ_i. Letting $\underset{\sim}{\lambda}$ denote the vector of mortality rates, the joint density for the λ_i given $\alpha, \underset{\sim}{\beta}$ is

$$\pi(\underset{\sim}{\lambda} \mid \alpha, \underset{\sim}{\beta}) = \prod_{i=1}^{\ell} \left\{ \frac{(\alpha e^{-\underset{\sim}{x}_i'\underset{\sim}{\beta}})^{\alpha} \lambda_i^{\alpha-1} \exp\{-\alpha e^{-\underset{\sim}{x}_i'\underset{\sim}{\beta}} \lambda_i\}}{\Gamma(\alpha)} \right\}.$$

We wish to maximize $\pi(\underset{\sim}{\lambda} \mid \alpha, \underset{\sim}{\beta})$ after incorporating the uncertainty in α and $\underset{\sim}{\beta}$ to obtain the posterior modal map.

This model is attractive because of the conjugacy in which the conditional posterior density of the λ_i is the simple gamma distribution. This permits us to construct Rao-Blackwellized estimators of the λ_i. Such an estimator has smaller mean square error than its empirical counterpart (Gelfand and Smith 1990). This makes it convenient to construct the posterior modal map. In the standard generalized linear model in which the $\log(\lambda_i)$ follow a normal linear model, it is not possible to obtain simple Rao-Blackwellized estimators of the λ_i; only empirical estimators can be easily obtained.

We take the shrinkage prior as the prior density for α

$$\pi(\alpha) = 1/(1+\alpha)^2, \ \alpha > 0.$$

One might prefer $\pi(\alpha) = a_0/(a_0+\alpha)^2, \ \alpha > 0$ where a_0 is the prior median of α, but we have found that inference is nonsensitive to the choice of a_0. We use a multivariate normal density as the prior for $\underset{\sim}{\beta}$

$$\underset{\sim}{\beta} \sim \text{Normal}(\underset{\sim}{\mu}_0, \kappa_0\Delta_0),$$

where $\underset{\sim}{\mu}_0$, Δ_0 and κ_0 (i.e., a variance inflation factor), are to be specified. We show how to specify $\underset{\sim}{\mu}_0$ and Δ_0 using a weighted least squares analysis in

Appendix A. Christiansen and Morris (1997) use a prior density of the form $\pi(\alpha) = \frac{a_0}{(a_0+\alpha)^2}$, but their prior specification for $\underset{\sim}{\beta}$ is noninformative (i.e., a flat prior).

Then the joint posterior distribution of all the parameters given $\underset{\sim}{d}$ is

$$\pi(\underset{\sim}{\lambda}, \alpha, \underset{\sim}{\beta} \mid \underset{\sim}{d}) \propto \prod_{i=1}^{\ell} \left\{ \frac{(n_i\lambda_i)^{d_i} e^{-n_i\lambda_i} (\alpha e^{-\underset{\sim}{x}_i'\underset{\sim}{\beta}})^{\alpha} \lambda_i^{\alpha-1} \exp\{-\alpha e^{-\underset{\sim}{x}_i'\underset{\sim}{\beta}} \lambda_i\}}{\Gamma(\alpha)} \right\}$$

$$\times \frac{e^{-\frac{1}{2}(\underset{\sim}{\beta}-\underset{\sim}{\mu}_0)'\Delta_0^{-1}(\underset{\sim}{\beta}-\underset{\sim}{\mu}_0)}}{(1+\alpha)^2}. \tag{2.3}$$

We note that this joint posterior density is a conditional density, but within the Bayesian paradigm, we do not call it a *conditional* posterior density. However, if in addition to the data, there is further conditioning on one or more parameters, Bayesians call this conditional density a conditional posterior density.

In Christiansen and Morris (1997) PRIMM is used to evaluate (2.3). Our method for constructing the posterior modal map requires a sampling-based method. So we use the Metropolis-Hastings sampler to fit the model; see Chib and Greenberg (1995) for a pedagogical discussion. We used the diagnostics reviewed by Cowles and Carlin (1996) to study convergence (i.e., we used the trace plots and autocorrelations) and we used the suggestion of Gelman, Roberts and Gilks (1996) to monitor the jumping probability in each Metropolis step. The jumping probability is obtained by counting the number of times the Markov chain moves from one state to another divided by the number of iterations after convergence; Gelman, Roberts and Gilks (1996) suggest that the jumping probability should be between .25 and .50.

To run the Metropolis-Hastings sampler, we just need the conditional posterior density of the λ_i, α and $\underset{\sim}{\beta}$. The condition posterior density for the λ_i is simple, and it is convenient to record that

$$\lambda_i \mid \alpha, \underset{\sim}{\beta}, d_i \sim \text{Gamma}(\alpha + d_i, n_i + \alpha e^{-\underset{\sim}{x}_i'\underset{\sim}{\beta}}) \tag{2.4}$$

and the conditional posterior density for α and $\underset{\sim}{\beta}$ is

$$\pi(\alpha, \underset{\sim}{\beta} \mid \underset{\sim}{\lambda}, \underset{\sim}{d}) \propto$$

$$\frac{e^{-\frac{1}{2}(\underset{\sim}{\beta}-\underset{\sim}{\mu}_0)'\Delta_0^{-1}(\underset{\sim}{\beta}-\underset{\sim}{\mu}_0)}}{(1+\alpha)^2} \prod_{i=1}^{\ell} \frac{(\alpha e^{-\underset{\sim}{x}_i'\underset{\sim}{\beta}})^{\alpha} \lambda_i^{\alpha-1} \exp\{-\alpha e^{-\underset{\sim}{x}_i'\underset{\sim}{\beta}} \lambda_i\}}{\Gamma(\alpha)}. \tag{2.5}$$

We draw α and $\underset{\sim}{\beta}$ simultaneously from the joint conditional posterior density using a Metropolis step with an independence chain. We obtain a proposal density for the Metropolis step by approximating $\pi(\alpha, \underset{\sim}{\beta} \mid \underset{\sim}{\lambda}, \underset{\sim}{d})$ by $\pi_a(\alpha, \underset{\sim}{\beta} \mid \underset{\sim}{\lambda}, \underset{\sim}{d})$ where

$$\pi_a(\alpha, \underset{\sim}{\beta} \mid \underset{\sim}{\lambda}, \underset{\sim}{d}) = g_1(\alpha \mid \underset{\sim}{\beta}, \underset{\sim}{\lambda}, \underset{\sim}{d}) g_2(\underset{\sim}{\beta} \mid \underset{\sim}{\lambda}, \underset{\sim}{d}),$$

$g_1(\alpha \mid \underset{\sim}{\beta}, \underset{\sim}{\lambda}, \underset{\sim}{d})$ and $g_2(\underset{\sim}{\beta} \mid \underset{\sim}{\lambda}, \underset{\sim}{d})$ are given in Appendix B. The Metropolis ratio is $\psi(\alpha, \underset{\sim}{\beta}) = \pi(\alpha, \underset{\sim}{\beta} \mid \underset{\sim}{\lambda}, \underset{\sim}{d}) / \pi_a(\alpha, \underset{\sim}{\beta} \mid \underset{\sim}{\lambda}, \underset{\sim}{d})$.

To compute our maps, we first need a random sample from the joint posterior density of $\Omega = (\alpha, \underset{\sim}{\beta})$. We obtain a random sample $\Omega^{(h)}, h = 1, \ldots, M$ ($M = 1000$) from the Metropolis-Hastings sampler. We ran the Metropolis-Hastings sampler for 5500 iterations, and we used a "burn in" of 500 iterations. Then, we picked every 5th from the remaining 5000 to make the autocorrelations among the iterates negligible. A further check on the jumping rate of the Metropolis-Hastings sampler shows the jumping probability is around 0.40 for all our activities. Also all the autocorrelations and numerical standard errors are small enough. Tuning of the Metropolis step is obtained by varying the parameter κ_1; see Appendix B. We found that $\kappa_1 = 1.50$ works fine.

Table 1: Sensitivity of inference about the mortality rates to the specification of κ_0: Comparison of Average (AVG) and standard deviation (STD) of posterior means (PM) and standard deviations (PSD) of the mortality rates for the 798 HSAs by κ_0

	PM		PSD	
κ_0	AVG	STD	AVG	STD
10	38.713	6.090	2.876	1.084
100	38.713	6.089	2.876	1.084
500	38.714	6.090	2.876	1.084
1000	38.713	6.089	2.875	1.084
10000	38.714	6.089	2.875	1.084
100000	38.714	6.090	2.876	1.084

Note: AVG and STD must be multiplied by 10^{-4}; κ_0 is the variance inflation factor.

We have specified the values of $\underset{\sim}{\mu}_0$ and Δ_0 in our analysis, and therefore a sensitivity analysis, which we studied through the variance inflation factor κ_0, is relevant. For various large values of κ_0 we have computed the posterior mean (PM) and posterior standard deviation (PSD) for the 798 mortality rates. Then, we took the average (AVG) and the standard deviation (STD) of the 798 values of the PMs and the PSDs respectively. The results are presented in Table 1.

For the six values of κ_0 from 10 to 100,000, there are virtually no changes. The results indicated that we can actually use noninformative priors (see Christiansen and Morris 1997) for a condition about propriety of the posterior density which is automatic in our model. In our empirical work we set $\kappa_0 = 10,000$ (i.e., essentially a noninformative prior).

An alternative Metropolis-Hastings sampler can be obtained. We have integrated out the λ_i to obtain the joint posterior density of α and $\underset{\sim}{\beta}$, and applied a procedure similar to the one for the conditional posterior density of α and $\underset{\sim}{\beta}$ in our current sampler. One can see that this procedure would save a little time in computation (i.e., the λ_i are drawn only in the output stage). Unfortunately, it was difficult to tune this version of the Metropolis-Hastings sampler (i.e., high correlations persist and we could not get autocorrelations down). It is possible to use a resampling method (not Markov chain Monte Carlo) to fit the model here, but we did not explore it.

Finally, we consider a measure, based on standardized cross-validation residuals, to assess the fit of the model. Let $\underset{\sim}{d}_{(i)}$ denote the set of all the d_i except d_i itself. Then letting $r_i = d_i/n_i$, we define the cross-validation residual as $\text{ARES}_i = r_i - E(r_i \mid \underset{\sim}{d}_{(i)})$, and the standardized cross-validation residual as

$$\text{DRES}_i = \text{ARES}_i/\text{SD}(r_i|\underset{\sim}{d}_{(i)}). \tag{2.6}$$

That is, the ith observed r_i is "held out" and compared with its point estimator, $E(r_i|\underset{\sim}{d}_{(i)})$, which is evaluated without using the observed d_i. We use the cross-validation residuals as a measure of concordance of the data with the model. In Figure 1 we have presented residual plots. Figure 1 (a) DRES versus predicted value shows that the Poisson-gamma regression model fits reasonable well with few possible outliers. Figure 1 (b) ARES versus standard deviation has bands at ARES $\pm$ 2SD and the points are mostly within these bands (see Nandram, Sedransk and Pickle 1999), indicating again that the Poisson-gamma regression model provides a good fit to the COPD mortality data for white males 65+.

3. Construction of the Posterior Modal Map

Our objective in this section is to show how to construct the posterior modal map. But we also show how to construct the posterior mean map for comparison.

Note that the posterior density of $\underset{\sim}{\lambda}$ is

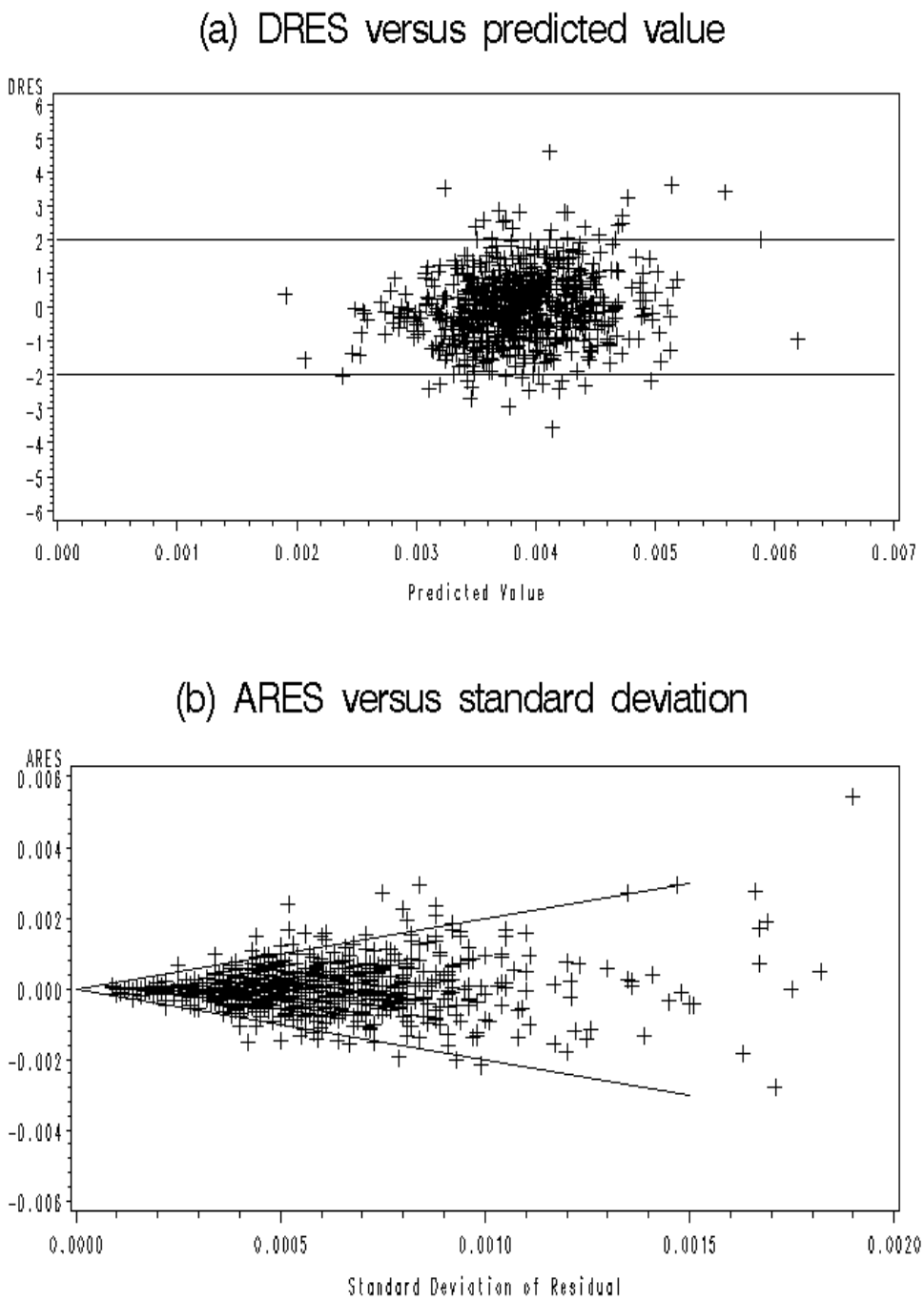

Figure 1: Residual plots of (a) DRES versus predicted value (b) ARES versus standard deviation.

$$\pi(\underset{\sim}{\lambda} \mid \underset{\sim}{d}) = \int_{\Omega} \pi(\underset{\sim}{\lambda} \mid \underset{\sim}{d}, \Omega)\pi(\Omega \mid \underset{\sim}{d})d\Omega \tag{3.1}$$

where the conditional posterior density of $\pi(\underset{\sim}{\lambda} \mid \underset{\sim}{d}, \Omega) = \prod_{i=1}^{\ell} \pi(\lambda_i \mid d_i, \Omega)$ and

$$\lambda_i \mid d_i, \Omega \overset{ind}{\sim} \text{Gamma}\left(\alpha + d_i, \quad n_i + \alpha e^{-\underset{\sim}{x}_i'\underset{\sim}{\beta}}\right). \tag{3.2}$$

Here Ω is $p+1$ dimensional vector, not too large, but $\underset{\sim}{\lambda}$ is a ℓ dimensional vector (i.e., $\ell = 798$, very large). As described in Section 2 we have a random sample $\underset{\sim}{\lambda}^{(1)}, \ldots, \underset{\sim}{\lambda}^{(M)}$ from $\pi(\underset{\sim}{\lambda} \mid \underset{\sim}{d})$. We now need to find the point $\underset{\sim}{\lambda}^*$ that maximizes $\pi(\underset{\sim}{\lambda} \mid \underset{\sim}{d})$ in (3.1). Note that it is impossible to maximize $\pi(\underset{\sim}{\lambda} \mid \underset{\sim}{d})$ in (3.1) directly (i.e., we need to integrate over the posterior density of Ω, $\pi(\Omega \mid \underset{\sim}{d})$ which does not exist in closed form). So that we can not use standard analytical optimization techniques (e.g., calculus). Our procedure calculates $\pi(\underset{\sim}{\lambda} \mid \underset{\sim}{d})$ at each of the points $\underset{\sim}{\lambda}^{(1)}, \ldots, \underset{\sim}{\lambda}^{(M)}$, and chooses the point that gives the largest $\pi(\underset{\sim}{\lambda} \mid \underset{\sim}{d})$. Thus the problem simply reduces to evaluating $\pi(\underset{\sim}{\lambda} \mid \underset{\sim}{d})$ at each of the points $\underset{\sim}{\lambda}^{(1)}, \ldots, \underset{\sim}{\lambda}^{(M)}$.

But first we show how to construct the posterior mean map using Rao-Blackwellized estimators for the λ_i. It is desirable to find these Rao-Blackwellized estimators because they have the smallest mean squared error (see Gelfand and Smith 1990). Letting $r_i = d_i/n_i,\ i = 1, \ldots, \ell$ denote the observed mortality rate and $\Lambda_i = n_i/(n_i + \alpha e^{-\underset{\sim}{x}_i'\underset{\sim}{\beta}})$, using (3.2) the conditional posterior mean of λ_i is

$$E(\lambda_i \mid \alpha, \underset{\sim}{\beta}, d_i) = \Lambda_i r_i + (1 - \Lambda_i)e^{\underset{\sim}{x}_i'\underset{\sim}{\beta}}.$$

As expected, this is a weighted average of the observed mortality rate and the prior mortality rate. It follows that the posterior mean (unconditional) of λ_i is

$$\mu_i = E(\lambda_i \mid \underset{\sim}{d}) = E_{\Omega|\underset{\sim}{d}}\{\Lambda_i r_i + (1 - \Lambda_i)e^{\underset{\sim}{x}_i'\underset{\sim}{\beta}}\}.$$

Note that because of the conditioning (posterior) on the data, μ_i is a function of the data. The Rao-Blackwellized estimator of μ_i is

$$\hat{\mu}_i = M^{-1}\sum_{h=1}^{M}\{\Lambda_i^{(h)} r_i + (1 - \Lambda_i^{(h)})e^{\underset{\sim}{x}_i'\underset{\sim}{\beta}^{(h)}}\} \tag{3.3}$$

where $\Lambda_i^{(h)} = n_i/(n_i + \alpha^{(h)} e^{-\underset{\sim}{x}_i'\underset{\sim}{\beta}^{(h)}})$ and $\Omega^{(h)} = (\alpha^{(h)}, \underset{\sim}{\beta}^{(h)}),\ h = 1, \ldots, M$ are the M iterates obtained from the Metropolis-Hastings sampler. The posterior mean map is obtained by mapping the $\hat{\mu}_i$ in (3.3) for all 798 HSAs.

The method for constructing the posterior modal map is computationally intensive, but it follows easily from the output of the Metropolis-Hastings sampler

already described. Again letting $\underset{\sim}{\lambda}$ denote the vector of all 798 λ_i, we need the mode of the joint posterior density, $\pi(\underset{\sim}{\lambda} \mid \underset{\sim}{d})$. Naturally, this is a very complex optimization problem because there are 798 variables, and $\pi(\underset{\sim}{\lambda} \mid \underset{\sim}{d})$ does not exist in closed form. Fortunately, we do not need to optimize $\pi(\underset{\sim}{\lambda} \mid \underset{\sim}{d})$ directly.

The procedure is to obtain the value of the posterior density $\pi(\underset{\sim}{\lambda} \mid \underset{\sim}{d})$ at each of the $M = 1000$ iterates $\underset{\sim}{\lambda}^{(h)},\ h = 1, \ldots, M$ obtained from the Metropolis-Hastings sampler (i.e., at $\lambda^{(h)}$ we compute $\pi(\underset{\sim}{\lambda}^{(h)} \mid \underset{\sim}{d})$, the ordinate of the posterior density at $\lambda^{(h)}$ which we denote by $C_h,\ h = 1, \ldots, M$). The posterior modal map consists of the values of $\underset{\sim}{\lambda}^{(h^*)}$ that maximize C_h among the M values $\underset{\sim}{\lambda}^{(h)},\ h = 1, \ldots, M$. Note that for this procedure we only use the iterates $(\underset{\sim}{\lambda}^{(h)}, \Omega^{(h)}),\ h = 1, \ldots, M$ and further sampling is not required.

Thus, we describe how to obtain the posterior ordinate c_h at $\underset{\sim}{\lambda}^{(h)}$ for any $h,\ h = 1, \ldots, M$. First, we note that

$$c_h = \pi(\underset{\sim}{\lambda}^{(h)} \mid \underset{\sim}{d}) = E_{\Omega|\underset{\sim}{d}}(\pi(\underset{\sim}{\lambda}^{(h)} \mid \underset{\sim}{d}, \Omega))$$

and a Rao-Blackwellized estimator of c_h is

$$\hat{c}_h = M^{-1} \sum_{k=1}^{M} \pi(\underset{\sim}{\lambda}^{(h)} | \underset{\sim}{d}, \Omega^{(k)}) \tag{3.4}$$

where $\pi(\underset{\sim}{\lambda}^{(h)} \mid \underset{\sim}{d}, \Omega^{(k)}) = \prod_{i=1}^{\ell} \pi(\lambda_i^{(h)} \mid d_i, \Omega^{(k)})$ and

$$\pi(\lambda_i^{(h)} \mid \underset{\sim}{d}, \Omega^{(k)}) =$$

$$\{n_i + \alpha^{(k)} e^{-\underset{\sim}{x}_i' \underset{\sim}{\beta}^{(k)}}\}^{d_i + \alpha^{(k)}} [\lambda_i^{(h) d_i + \alpha^{(k)} - 1} \exp\{-(n_i + \alpha^{(k)} e^{-\underset{\sim}{x}_i' \underset{\sim}{\beta}^{(k)}}) \lambda_i^{(h)})\}] / \Gamma(d_i + \alpha^{(k)}).$$

Thus, the algorithm for constructing the posterior modal map has the following three steps.

(a) Perform the Metropolis-Hastings sampler to obtain M iterates $(\underset{\sim}{\lambda}^{(h)}, \Omega^{(h)})$, $h = 1, \ldots, M$.

(b) Compute the ordinate $\hat{c}_h$ of the posterior density at each of $h = 1, \ldots, M$ iterates.

(c) Sort the $\hat{c}_h,\ h = 1, \ldots, M$ in increasing order. The posterior modal map is obtained by taking $\underset{\sim}{\lambda}^{(h^*)}$, where h^* corresponds to the maximum of the c_h.

To make inference, (a) has to be performed anyway. Both (b) and (c) are easy to perform.

For the COPD data in our procedure the joint mode has roughly 382 HSAs with modes bigger than means and 416 have modes smaller than means. Our

objective in the modal map is to provide the map of the mortality rates corresponding to the coordinates of the point where the joint posterior density of the mortality rates is the highest. For highly dispersed and sparse data, the posterior distribution of the rate is often skewed. Therefore, the posterior mean can be different from the posterior mode, leading to an inaccurate representation by the posterior mean map.

4. Data Analysis and Map Comparison

In this section we analyze the COPD data, white males age 65+, using the Poisson regression model. More importantly, we compare the posterior modal map and the posterior mean map.

In Table 2 we present the posterior mean (PM), posterior standard deviation (PSD) numerical standard error (NSE) and the 95% credible intervals for the regression coefficients. First, the NSEs are very small indicating that the computation is doing well, and thus the results are trustworthy. The PSDs are all small when compared with the PMs.

Table 2: Posterior means (PM), standard deviations (PSD), numerical standard errors (NSE) and 95% credible intervals for the regression coefficients

Covariate	PM	PSD	NSE	Interval
Intercept	-5.948	0.072	0.014	(-6.088, -5.781)
Lung cancer rate	0.008	0.001	0.000	(0.006, 0.010)
Population Density	-0.497	0.108	0.022	(-0.726, -0.278)
Elevation	0.004	0.001	0.000	(0.002, 0.005)
Annual rainfall	-0.401	0.101	0.021	(-0.593, -0.160)

NOTE: The NSEs are obtained using the batch means method with batches of length 25 from the output sample of 1000 iterates.

Lung cancer rate and elevation have positive effects on mortality. HSAs where more people smoke tend to have a higher COPD mortality (see Morris and Munasinghe 1994) and extreme climatic conditions aggravate existing asthma and bronchitis (Bates 1989), as is living at high altitudes because of the reduced oxygen supply (Schoene 1999).

Population density and annual rainfall have negative effects on mortality. Those places with a high population density usually have better medical services,

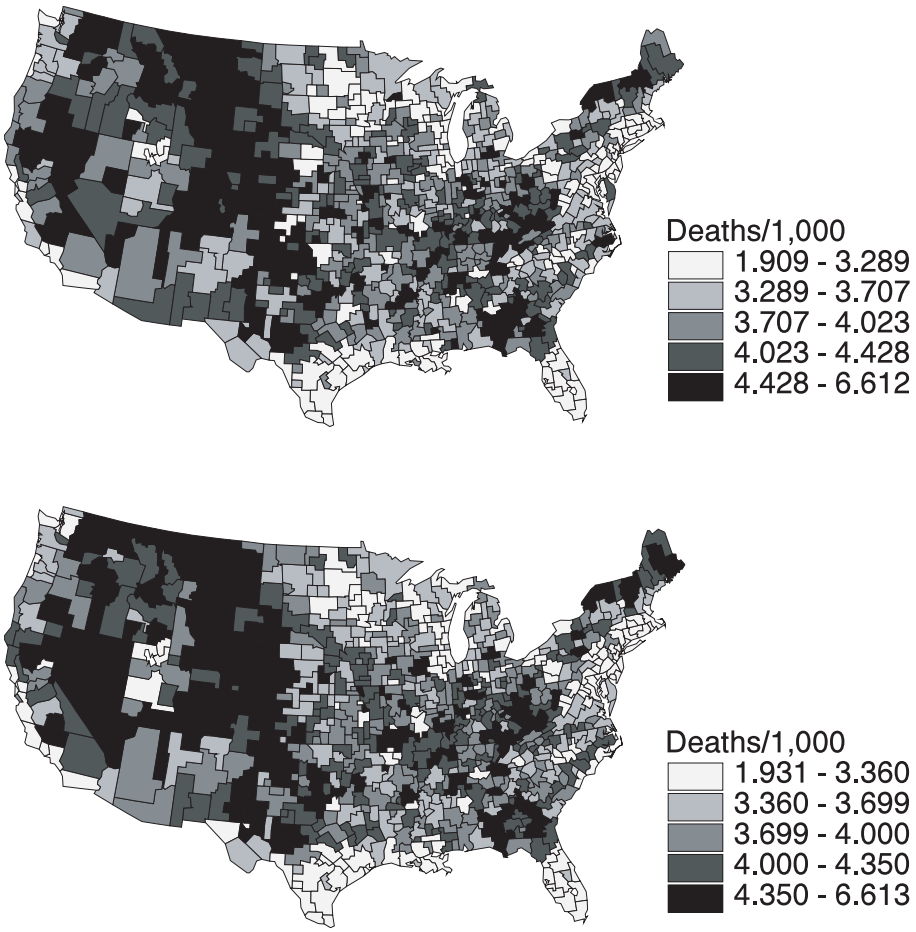

Figure 2: A comparison of the posterior mean and modal maps for COPD, white males age 65+, top: Posterior modal map; bottom: Posterior mean map.

and when there is an emergency, people living in a remote area are more likely to be delayed by the long travel to the nearest hospital (see Nandram, Sedransk and Pickle 1999). Repeated exposure to particulate matter and other air pollutants, primarily from traffic exhaust and coal-burning power plants, can aggravate existing lung conditions and can even cause death (English, Neutra, Scalf, Sullivan Waller and Zhu 1999, Sunyer, Schwartz, Tobias, Macfarlane, Garcia and Anto 2000). In particular, small airborne particles such as SO_2 found in urban air pollution can be deposited deep in the lungs, causing severe pulmonary effects (Schwartz and Neas 2000 and Sunyer, Schwartz, Tobias, Macfarlane, Garcia and Anto 2000). Aerosolized toxins and viruses can be inhaled in dusty environments, causing pulmonary effects (National Center for Health Statistics). Rainfall, on the contrary, can lower the density of airborne particles and dust in the air, thus

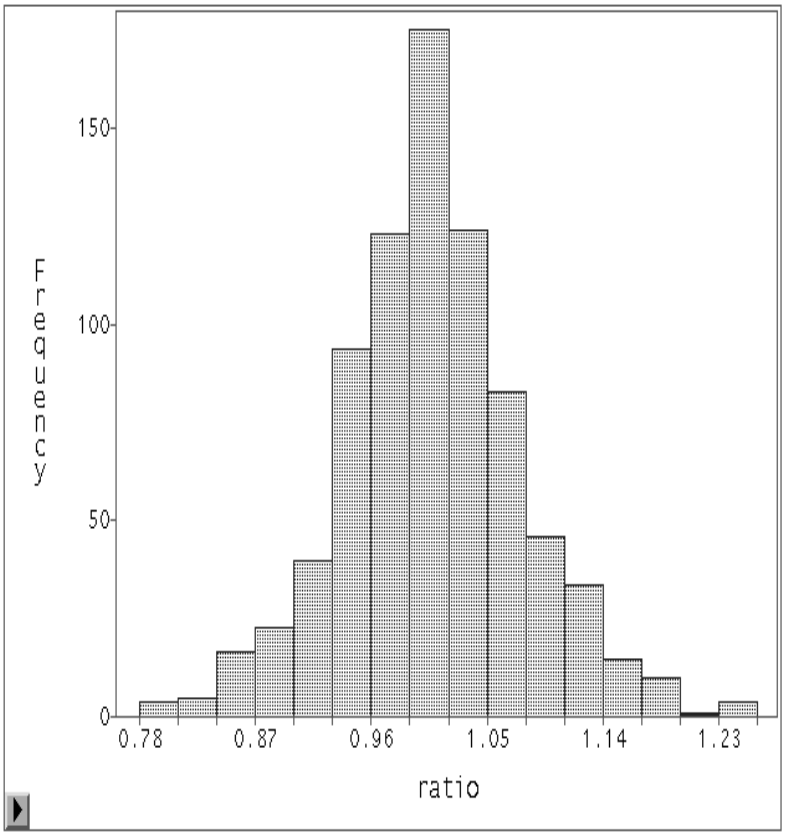

Figure 3: Histogram of the ratio of the posterior mean tates to posterior modal rates for the 798 HSAs

lower the chance of contracting a pulmonary disease.

Next, we compare the posterior modal map and the posterior mean map. In Figure 2 we present the choropleth maps based on the mean quintiles in which the mortality rate is per 1,000 white males 65+. The two maps are mostly similar especially on the eastern half of the United states. Most of the differences are noticeable on the western half of the United states. We have studied the differences in these two maps in greater detail.

First, we compute the ratio of $\hat{\mu}_i$ in (3.3) to $\lambda^{(h^*)}$, the joint modal rates obtained from our algorithm, namely

$$R_i = \hat{\mu}_i / \lambda_i^{(h^*)}.$$

We have presented the distribution of the ratios for the 798 HSAs in Figure

3. The distribution is roughly symmetric and the five number summaries are 0.80, 0.97, 1.01, 1.05, 1.26.

We have also studied the coefficient of skewness for each of the λ_i for the 798 HSAs (i.e., for each HSA the iterates from the Metropolis-Hastings sampler are used to estimate the coefficient of skewness). Then we drew the histogram of the 798 sample coefficients: the five number summaries are $-0.16, 0.08, 0.15, 0.21, 0.45$; only 43 of the 798 HSAs have negative skewness. In fact, the individual posterior modal and posterior mean of the λ_i are very close. We have drawn the map corresponding to the individual modes, and we have found virtually no difference when compared with the posterior mean map. This is expected because the 95% credible interval for α is (19.49, 36.54); thereby making the posterior means approximately the same as the individual posterior modes. As we have pointed out, the problem of mapping the individual posterior mode is not our objective, and mapping the overall posterior mode should be the objective of a scientist who uses the Bayesian paradigm.

Table 3: Cross-tabulation of the 798 HSAs by modal and mean quintiles of the mortality rates

	Mode				
Mean	1.91 − 3.29	3.29 − 3.71	3.71 − 4.02	4.02 − 4.43	4.43 − 6.61
1.91 − 3.36	133	24	3	0	0
3.36 − 3.70	24	91	38	7	0
3.70 − 4.00	2	38	68	45	7
4.00 − 4.35	1	6	48	75	30
4.35 − 6.61	0	1	3	33	121

In Table 3 we have cross classified the 798 HSAs according to which quintile they belong to in the mean map and the modal map. It is good that many of the HSAs lie along the diagonal of the 5×5 table. But there is a substantial number of the HSAs off the diagonal. Of the 160 HSAs in the first mean quintile, there are 24 HSAs in the second modal quintile, and of the 158 in the fifth mean quintile, there are 33 in the fourth modal quintile. There are greater changes for the second, third and fourth quintiles. For example, for the third quintile (i.e., the middle one) there are 68 in the third modal quintile and 92 off diagonal.

In Table 4 we have presented some specific HSAs in which there are large differences between the posterior modal map and the posterior mean map. The HSA 480 (Chicot and Ashley counties in Arkansas) is in the first quintile (pink) in the mean map but is in third quintile in the modal map (middle red). The HSA 575 (Finney and Scott counties in Kansas) is in the fourth quintile (second top red) in the mean map but only in the first quintile (pink) in the modal map.

It is clear that while there are similarities between the modal map and the mean map, there are important differences in some HSAs as well. Thus, it will be beneficial to construct the posterior modal map in disease mapping because it is the most likely representation of the mortality rates as discussed before.

Table 4: Examples of HSAs which have very different quintile classification in the modal and mean maps

Quintile			
Mean	Mode	HSA Code	HSA Name
1	3	480	Chicot and Ashley, AR
		495	Denton and Cooke, TX
		662	Clay, KS
2	4	3	Sussex, DE and Wicomico, MD
		605	Stark and Golden Valley, ND
		613	Red Willow, NE and Decatur, KS
		642	Cloud and Republic, KS
3	5	471	Jackson (Pascagoula) and George, MS
		481	Hale and Floyd, TX
		469	Woodward and Ellis, OK
		399	Jackson, WI
		661	Harper and Kingman, KS
		669	Barber, KS
		607	Henry and St. Clair, MO
4	1	575	Finney and Scott, KS
4	2	356	Williams and Defiance, OH
3	1	524	Collingsworth, TX

NOTE: As an example, HSAs 480, 490, 662 fall in the first quintile for the mean map and in the third quintile for the modal map.

5. Concluding Remarks

We have shown how to (a) fit the model of Christiansen and Morris (1997) (b) construct the posterior mean map and (c) construct the posterior modal map. In fact (c) is our key contribution. We obtain (c) using an output analysis

from the Metropolis-Hastings sampler and (b) is done in order to compare the posterior modal map with the posterior mean map. We have shown that there are differences between the posterior mean map and the posterior modal map. One example, the HSA 575 , consisting of the Finney and Scott counties in Kansas, is in the fourth quintile in the mean map, but only in the first quintile in the modal map.

We make one remark. A possible posterior modal map can be obtained by finding the posterior modal rate for each area. Then these can be mapped for all areas to provide a choropleth map. However, this has not been our intention because from the Bayesian point of view this is not the map with the highest posterior probability. Among the set of 1000 maps obtained from Metropolis-Hastings sampler our procedure finds the one that maximizes the joint posterior density over the 798 health service areas. We have repeated our procedure with 10,000 maps, and we have found minor changes. We believe that this procedure of finding the posterior modal map is novel.

The simple Poisson regression model has wide applicability for rare events. The conditional conjugacy in the Poisson-gamma regression model provides some simplification in our analysis. Each λ_i has a gamma conditional posterior density, and they are independent. This helps in finding the mode of the joint posterior density. With nonconjugacy, (e.g., a Poisson-normal model) there will be difficulty in finding the mode of the joint posterior density. The models of Nandram, Sedransk and Pickle (1999, 2000) and Waller, Carlin, Xia and Gelfand (1997) are of the Poisson-normal type. The Poisson-gamma regression is relatively robust, so that one should not bother too much with the conjugacy of the Poisson sampling process and the gamma prior distribution. It is possible to add more sources of variation and more stages in the hierarchical Poisson-gamma model. Finding the posterior modal map in both the Poisson-gamma and Poisson-normal models in more complex problems needs further research.

Appendix A: Specification of $\underset{\sim}{\mu}_0$ and Δ_0

Letting $\tilde{\lambda}_i = d_i/n_i$, an estimator of λ_i is

$$\hat{\lambda}_i = \begin{cases} \tilde{\lambda}_i, & d_i > 0 \\ \bar{d}/\bar{n}, & d_i = 0, \end{cases}$$

where $\bar{n} = \sum_{i=1}^{\ell} n_i/\ell$ and $\bar{d} = \sum_{i=1}^{\ell} d_i/\ell$. By the Poisson assumption, given λ_i,

$$E\{\log(\tilde{\lambda}_i)\} \approx \lambda_i \text{ and } Var\{\log(\tilde{\lambda}_i)\} \approx 1/(n_i\lambda_i).$$

Using the prior density for the λ_i, $E(\lambda_i \mid \alpha, \underset{\sim}{\beta}) = e^{\underset{\sim}{x}_i'\underset{\sim}{\beta}}$ and $E\{\log(\lambda_i) \mid \alpha, \underset{\sim}{\beta}\} \approx$

$\underset{\sim}{x}_i'\underset{\sim}{\beta}$. Thus, we assume that

$$\log(\hat{\lambda}_i) = \underset{\sim}{x}_i'\underset{\sim}{\beta} + e_i, \quad e_i \overset{ind}{\sim} \text{Normal}\{0,\ \gamma^2/(n_i\hat{\lambda}_i)\}, \quad i = 1, \ldots, \ell, \tag{A.1}$$

where γ^2 is an unknown scale factor.

We compute weighted least square estimators in (A.1). Let $\underset{\sim}{Y} = (\log\hat{\lambda}_1, \ldots, \log\hat{\lambda}_\ell)'$, $X = (\underset{\sim}{x}_1, \ldots, \underset{\sim}{x}_\ell)'$ and $W = \text{diagonal}(n_1\hat{\lambda}_1, \ldots, n_\ell\hat{\lambda}_\ell)$. Then, the least square estimator of $\underset{\sim}{\beta}$ is

$$\hat{\underset{\sim}{\beta}} = (X'W^{-1}X)^{-1}(X'W^{-1}Y) \tag{A.2}$$

and $Cov(\hat{\underset{\sim}{\beta}})$ is estimated by

$$\widehat{Cov}(\hat{\underset{\sim}{\beta}}) = (X'W^{-1}X)^{-1}\hat{\gamma}^2 \tag{A.3}$$

where $\hat{\gamma}^2 = (Y - X\hat{\underset{\sim}{\beta}})'W^{-1}(Y - X\hat{\underset{\sim}{\beta}})/(n-p)$.

Finally, we specify $\underset{\sim}{\mu}_0$ and Δ_0 by taking $\underset{\sim}{\mu}_0 = \hat{\underset{\sim}{\beta}}$ in (A.2) and $\Delta_0 = \kappa_0\widehat{Cov}(\hat{\underset{\sim}{\beta}})$ in (A.3), where κ_0 is a variance inflation factor. By experimentation, we choose κ_0 to be large so that the prior density for $\underset{\sim}{\beta}$ is proper and barely informative. For our data analysis, after a sensitivity analysis, which reveals nonsensitivity to κ_0, we choose $\kappa_0 = 10000$.

Appendix B: A multivariate proposal density for $(\alpha, \underset{\sim}{\beta})$

We obtain a proposal density for the conditional posterior density of $(\alpha, \underset{\sim}{\beta})$ using the multivariate normal density in which the mean is taken to be the mode and the variance is the negative inverse Hessian matrix. Taking the logarithm of the conditional posterior density, we have

$$\Delta(\alpha, \underset{\sim}{\beta}) = \sum_{i=1}^{\ell}\left\{(\alpha-1)\log\lambda_i - \alpha e^{-\underset{\sim}{x}_i'\underset{\sim}{\beta}}\lambda_i + \alpha(\log(\alpha) - \underset{\sim}{x}_i'\underset{\sim}{\beta}) - \log\Gamma(\alpha)\right\}$$
$$-2\log(1+\alpha) - \frac{1}{2}(\underset{\sim}{\beta} - \underset{\sim}{\mu}_0)'\Delta_0^{-1}(\underset{\sim}{\beta} - \underset{\sim}{\mu}_0). \tag{B.1}$$

We obtain the modal values, $\hat{\alpha}$ and $\hat{\underset{\sim}{\beta}}$, of α and $\underset{\sim}{\beta}$ in (B.1) using the Nelder-Mead algorithm. Thus, the mean of the conditional posterior density of $(\alpha, \underset{\sim}{\beta})$ is $(\hat{\alpha}, \hat{\underset{\sim}{\beta}})$. Thus, we next construct a surrogate for the variance using the Hessian matrix.

Letting $\psi'(\cdot)$ denote the trigamma function, the second derivative of $\Delta(\alpha, \underset{\sim}{\beta})$ with respect to α is

$$d = \frac{\partial^2\Delta}{\partial\alpha^2} = \sum_{i=1}^{\ell}\left\{1/\alpha - \psi'(\alpha)\right\} + 2/(1+\alpha)^2,$$

the second derivative, $H_{\underset{\sim}{\beta}}$, with respect to $\underset{\sim}{\beta}$ is

$$H_{\underset{\sim}{\beta}} = -[\Delta_0^{-1} + \alpha \sum_{i=1}^{\ell} \lambda_i e^{-\underset{\sim}{x}_i' \underset{\sim}{\beta}} \underset{\sim}{x}_i \underset{\sim}{x}_i']$$

and the second derivative with respect to both α and $\underset{\sim}{\beta}$ is

$$\underset{\sim}{c} = -\sum_{i=1}^{\ell} (1 - \lambda_i e^{-\underset{\sim}{x}_i' \underset{\sim}{\beta}}) \underset{\sim}{x}_i.$$

Then, an approximation for the covariance matrix of $(\alpha, \underset{\sim}{\beta})$ in the conditional posterior density is

$$\Sigma = \begin{bmatrix} \sigma_\alpha^2 & \underset{\sim}{\nu}' \\ \underset{\sim}{\nu} & \Delta_{\underset{\sim}{\beta}} \end{bmatrix} = -\kappa_1 \begin{bmatrix} d & \underset{\sim}{c}' \\ \underset{\sim}{c} & H_{\underset{\sim}{\beta}} \end{bmatrix}^{-1}, \tag{B.2}$$

where κ_1 is a tuning constant. We complete the process for the approximation by replacing $(\alpha, \underset{\sim}{\beta})$ in (B.2) by the modal estimates, $(\hat{\alpha}, \hat{\underset{\sim}{\beta}})$, to obtain $\hat{\Sigma}$ with components $\hat{\sigma}_{\hat{\alpha}}^2$, $\hat{\underset{\sim}{\nu}}$ and $\hat{\Delta}_{\hat{\underset{\sim}{\beta}}}$.

Finally, the multivariate proposal density is obtained by taking

$$\alpha \mid \underset{\sim}{\beta} \sim \text{Gamma}(a, b) \text{ and } \underset{\sim}{\beta} \sim \text{Normal}(\hat{\underset{\sim}{\beta}}, \hat{\Delta}_{\hat{\underset{\sim}{\beta}}})$$

with $a = \tilde{\mu}^2/\tilde{\sigma}^2$ and $b = \tilde{\mu}/\tilde{\sigma}^2$ where $\tilde{\mu} = \hat{\alpha} + \hat{\underset{\sim}{\nu}}' \hat{\Delta}_{\hat{\underset{\sim}{\beta}}}^{-1} (\underset{\sim}{\beta} - \hat{\underset{\sim}{\beta}})$ and $\tilde{\sigma}^2 = \hat{\sigma}_{\hat{\alpha}}^2 - \hat{\underset{\sim}{\nu}}' \hat{\Delta}_{\hat{\underset{\sim}{\beta}}}^{-1} \hat{\underset{\sim}{\nu}}$.

References

Bates, D. V. (1989). *Respiratory Function in Disease.* W. B. Saunders.

Christiansen, C. L. and Morris, C. N. (1997). Hierarchical Poisson regression modeling. *Journal of the American Statistical Association* **92**, 618-631.

Chib, S. and Greenberg, E. (1995). Understanding the Metropolis-Hastings algorithm. *The American Statistician* **49**, 327-335.

Cowles, M. and Carlin, B. (1996). Markov chain Monte Carlo diagnostics: A comparative review. *Journal of the American Statistical Association* **91**, 883-904.

English, P., Neutra, R., Scalf, R., Sullivan, M., Waller, L. and Zhu, L. (1999). Examining associations between childhood asthma and traffic flow using a geographic information system. *Environmental Health Perspectives* **107**, 761-767.

Gelfand, A. E. and Smith, A. F. M. (1990). Sampling-based approaches to calculate marginal densities. *Journal of the American Statistical Association* **85**, 398-409.

Gelman, A. and Roberts, G. O. and Gilks, W. R. (1996). Efficient Metropolis jumping rules. In *Bayesian Statistics* **5** (Edited by J. M. Bernardo, J. O. Berger, A. P. Dawid A. F. M. Smith), 599-607. Oxford University Press.

Morris, R. D. and Munasinghe, R. L. (1994). Geographic variability in hospital admission rates for respiratory disease among the elderly in the United States. *Chest* **106**, 1172-1181.

Nandram, B. (2000). Bayesian generalized linear models for inference about small areas. *Generalized Linear Models: A Bayesian Perspective* (Edited by D. K., Dey, S. K., Ghosh, and B. K., Mallick), 91-114. Marcel Dekker.

Nandram, B., Sedransk, J. and Pickle, L. W. (1999). Bayesian analysis of mortality rates for U.S. health service areas. *Sankhya B* **61**, 145-165.

Nandram, B., Sedransk, J. and Pickle, L. W. (2000). Bayesian analysis and mapping of mortality rates for chronic obstructive pulmonary disease. *Journal of the American Statistical Association* **95**, 1110-1118.

National Center for Health Statistics (1998). Health, United States. DHHS Publication Number (PHS) 98-1232, Hyattsville, MD: National Center for Health Statistics.

Robert, C. P. and Casella, G. (1999). *Monte Carlo Statistical Methods*. Springer.

Schwartz, J. and Neas, L. M. (2000). Fine particles are more strongly associated than coarse particles with acute respiratory health effects in school children. *Epidemiology* **11**, 6-10.

Sunyer, J., Schwartz, J., Tobias, A., Macfarlane, D., Garcia, J., and Anto, J. M. (2000). Patients with chronic obstructive pulmonary disease are at increased risk of death associated with urban particle air pollution: A case-crossover analysis. *American Journal of Epidemiology* **151**, 50-56.

Schoene, R. B. (1999). Lung disease at high altitude. *Advances in Experimental Medicine and Biology* **474**, 47-56.

Waller, L., Carlin, B., Xia, H. and Gelfand, A. E. (1997). Hierarchical spatio-temporal mapping of disease rates. *Journal of the American Statistical Association* **92**, 607-617.

Received November 11, 2003; accepted December 15, 2003.

Balgobin Nandram[1]
Office of Research and Methodology
National Center for Health Statistics
3311 Toledo Road
Hyattsville, MD 20782, USA
bbn9@cdc.gov

Jie Liu
306 Hill Building
Research Triangle Institute
3040 Cornwallis Road
RTP, NC 27709, USA
jliu@rti.org

Jai Won Choi
Office of Research and Methodology
National Center for Health Statistics
3311 Toledo Road
Hyattsville, MD 20782, USA
jwc7@cdc.gov

[1]Balgobin Nandram (balnan@wpi.edu) is on sabbatical leave from the Department of Mathematical Sciences, Worcester Polytechnic Institute, 100 Institute Road, Worcester, MA 01609-2280.