NISS

# Spatio-Temporal Interaction with Disease Mapping

Dongchu Sun, Robert K. Tsutakawa,
Hoon Kim, and Zhuoqiong He

Technical Report Number 98
April, 1999

National Institute of Statistical Sciences
19 T. W. Alexander Drive
PO Box 14006
Research Triangle Park, NC 27709-4006
www.niss.org

# Spatio-Temporal Interaction with Disease Mapping

Dongchu Sun, Robert K. Tsutakawa, Hoon Kim and Zhuoqiong He

SUMMARY

Markov chain Monte Carlo methods are used to estimate mortality rates under a Bayesian hierarchical model. Spatial correlations are introduced to examine spatial effects relative to both regional and regional changes over time by groups. A special feature of the models is the inclusion of longitudinal variables which will describe temporal trends in mortality or incidences for different population groups. Disease maps are used to illustrate the role of different parameters in the model and pinpointing areas of interesting patterns. The methods are demonstrated by male cancer mortality data from the state of Missouri during 1973–1992. Of special interest will be the geographic variations in the trend of lung cancer mortality over the recent past. Marginal posterior distributions are used to examine effects due to spatial correlations and age difference in temporal trends. Numerical results from the Missouri data show that though spatial correlations exist, they do not have a large effect on the estimated mortality rates.

KEY WORDS: Bayesian prediction; Gibbs sampling; Linear mixed models; Multivariate normal; Markov chain Monte Carlo; Mortality rates; Variance Components.

# 1 Introduction

Disease maps can be used to highlight geographic areas with high and low incidence or mortality rates of a specific diseases, such as lung cancer, and the variability of such rates over a state or country. They can also be used to detect spatial clusters which may be due to common environmental, demographic or cultural effects shared by neighboring regions. Although these maps are often constructed with countries or states as basic units, the boundaries between such units can be epidemiologically arbitrary when the neighboring units are homogeneous.

Mapping of crude rates can be non-informative or misleading when the sizes of the population for some of the units are small, resulting in large variability in the estimated rates, and making it difficult to distinguish chance variability from genuine differences. Pooling of neighboring units often masks important real differences, which could aid in pinpointing potential causes.

An example of the mapping of crude rates is illustrated in Figure 1, which shows annual male lung cancer mortality rates per 100,000 population by age group, county and period for the state of Missouri in central USA. The age groups are 45-54, 55-64, 65-74 and 75+. The periods are 1973-1977, 1978-1982, 1983-1987 and 1988-1992. The state has two large population centers, Kansas city and St. Louis, several mid-size cities and a large number of sparely populated rural communities. The city of St. Louis, which is not a part of any county, is quite urban, and differs from the surrounding St. Louis county environmentally and socio-economically. By including the city of St. Louis as a separated county, the state is divided into a total of 115 counties.

In Figure 1, we see the general increase in crude rates over the lower 3 age groups but no general increase from the third to the fourth age group. There are some signs of an increase over time for the two older groups. It is more difficult to detect counties or clusters of counties with high (or low) rates since the extreme rates tend not to appear consistently over county or time. Since the excessive rates typically occur in small counties, where a difference of a few deaths can have noticeable effect on the crude rates, it is important to separate the different sources of variations.

Recent developments in empirical Bayes and Bayesian hierarchical modeling has made it possible to obtain stable estimates for small areas, by using information from all of the areas to obtain estimates for individual areas. The development on this topic has been greatly enhanced by rapidly developing computational tools such as the Gibbs sampler (Gelfand and Smith[1]) and other Markov chain Monte Carlo (MCMC) methods (Tanner[2]) for posterior analysis.

The literature on disease mapping has also grown at a rapid pace. One of the earlier works is

by Breslow and Day[3] who assumed that for a homogeneous population of size n, the frequency of deaths $Y$, due to some disease over a given period, follows a Poisson distribution with mean np, where $p$ is the rate per individual and can be factored into several components such as age, sex, and region. They computed maximum likelihood estimates of $p$ for the different sub-populations. Empirical Bayes techniques for mortality rates were introduced by Manton, Woodbury, and Stallard[4], Tsutakawa, Shoop, and Marienfeld[5], and Tsutakawa[6]. Different types of spatial correlations among neighboring regions were proposed by Whittle[7], Besag[8], and Clayton and Kaldor[9]. The distinction between the simultaneously autoregressive model by Whittle and the conditionally autoregressive model by Clayton and Kaldor has been discussed by Cliff and Ord[10] and Cressie[11]. Clayton and Bernardinelli[12] pointed out that the conditionally autoregressive model in Clayton and Kaldor is not appropriate for irregular lattices since the conditional variance of the regional effect does not depend on the neighborhood structure. Various modifications and generalization with applications can be found in Besag, York, and Mollié[13], Marshall[14], Carlin and Louis[15] and Ghosh, Natarajan, Stroud, and Carlin[16], among others.

The effects of age, sex, and race on cancer mortality are well known. There is less known about how to analyze spatio-temporal effects. Lee and Lin[17] discuss temporal effects in the context of an age-period-cohort model by a Poisson mixed effect loglinear model using marginal maximum likelihood, however without geographic effects. Bernardinelli *et al.*[18,19] introduce spatio-temporal effects in terms of a Poisson loglinear model where the rate of change depends on the region, however without covariates. The analysis of mortality rates for cancer would not be very useful without the use of age-sex specific rates. Waller, Carlin, Xia, and Gelfand[20] introduced a spatio-temporal model where the spatial effects are nested within time. For disease such as cancer, one would expect the spatial effects to be systematically cross classified with time and not have a separate distribution within each time period. There is also some related work on Bayesian hierarchical age-period-cohort modeling in Besag, *et al.*[21], on the use of random covariates, e.g., amount of smoking, by Xia and Carlin[22] and on the use of time varying covariates by Knorr-Held and Besag[23].

Motivated by the Missouri data, topics of special interest in this paper will be the following.

a. Demographic effects (age will be used for illustration, though sex and race may also be added.)

b. Spatial effects among different geographic units.

c. Temporal effects on different demographic groups.

d. Regional changes over time.

e. Spatial correlation among regional changes over time.

f. Extra variation not explained by the mixed linear model.

g. Propriety of posterior distribution under noninformative prior distributions.

h. MCMC computing and Bayesian model fitting.

The extra variation in f has been considered by Tsutakawa[6] through a Poisson-gamma model, Besag *et al.*[13] in their spatial models, and Ghosh *et al.*[16] as a term in a generalized mixed linear model, among several others. There is a nice discussion of the interpretation of extra variation and spatially correlated random effects in space-time models in Knorr-Held and Besag[23]. Our spatial effects have two components, one to represent the overall difference among regions and another to represent the rate of change over time for the different regions.

In Section 2, we introduce a loglinear mixed model for mortality rates, where age effects are fixed and regional effects are random. A conditionally autoregressive model (CAR) is used to model the random effects. The model we use here is a modification of an intrinsic autoregressive model proposed by Besag *et al.*[13] and includes an additional parameter which indicates the amount of correlation among neighboring regions. Some properties of this model have been presented by Sun, Tsutakawa and Speckman[24]. The prior distributions for the variance components and spatial correlations are specified. In Section 3, we present conditions on the noninformative priors so that the posterior distributions are proper. In Section 4, estimation of the parameters via MCMC is proposed. The available conditional distributions are summarized. Numerical results on age effects, regional effects, variance components, spatial correlations are interpreted. It will be seen that these estimates are quite robust in terms of the choices of hyperparameters. The convergence of Gibbs sampling is investigated along the line of Gelman and Rubin[25] diagnostics. Finally, data analysis via disease mapping is presented in Section 5. Estimated mortality rates are plotted in the maps for each of the four age groups and four time periods. Various other plots are used to examine the regional features of the hierarchical model. Readers who are primarily interested in the model and numerical results may wish to skip the technical details presented in Sections 3 and 4.1.

# 2 Hierarchical Model for Disease Rates

## 2.1 Loglinear Model

For a given county, age group, and time period, we first assume that the frequency of deaths $Y$, due to a specific cause during a given time period, has a Poisson distribution with mean $np$, where $n$ is

the midperiod population size and p the rate per individual for this target population. Alternatively, we may consider using the binomial distribution since $n$ is known. However, in dealing with chronic diseases such as cancer, where relatively few deaths occur "randomly" over time within each age group, whose members vary throughout the period due to aging, the Poisson distribution is more appropriate than the binomial, which assumes a constant population of individuals. In order to represent the variability in $p$ due to county, age and time we model the rate by a loglinear model containing an extra variation term,

$$\log(p_{ijk}) = \theta_j + Z_i + (\mu_j + W_i)(t_k - \bar{t}) + e_{ijk}, \tag{1}$$

where $\bar{t} = K^{-1} \sum_{k=1}^{K} t_k$. Let $Y_{ijk}$ be the corresponding observed frequency of deaths for the ith county, jth age group and kth time period, $i = 1, \cdots, I; j = 1, \cdots, J; k = 1, \cdots, K$. For the data set considered in the paper, $(I, J, K) = (115, 4, 4)$. Thus $\theta_j$ represents the effect of the $jth$ age group, $Z_i$ the effect of the $ith$ county and $t_k$ the midpoint of the $kth$ time period. The change over time is represented by rate of change $(\mu_j + W_i)$ for the $jth$ age group in the $ith$ county. We treat $\theta_j$ and $\mu_j$ as fixed effects with some prior distribution and $Z_i$ and $W_i$ as random effects with some distribution whose hyperparameters will in turn have prior distributions. The extra variations $e_{ijk}$ are random effects, which have a prior distribution whose hyperparameters also have priors. These extra variations include other sources of variations not explained by the linear part and could include higher order interactions as well as unknown sources of variability in $p$. The non-uniqueness of the above parameterization will be eliminated through prior distributions to be discussed later. Conditionally on the rate $p_{ijk}$, we assume $Y_{ijk}$ are independent Poisson random variables with means $n_{ijk}p_{ijk}$, where $n_{ijk}$ are the population sizes.

This model incorporates a particular form of spatio-temporal interaction, namely, increase in log disease risk where the county slopes are allowed to have spatial correlation. Waller, Carlin and Xia[26] consider an extension of the Waller *et al.*[20] model where they include a random effect to each year, independent of the spatial effects. Their model allows a more general time trend than the linear one by considering an autoregressive prior for the temporal effects, but does not explicitly deal with space-time interaction.

## 2.2 Distributions of $Z_i$ and $W_i$

We represent the spatial correlation among counties by a modification of the intrinsically autoregressive model introduced by Besag *et al.* [13]. We first define the $I \times I$ adjacency matrix $\mathbf{C} = (C_{uv})$

where $C_{uv} = 1$ if two regions $u$ and $v$ share a common boundary, $C_{uv} = 0$ otherwise, with $C_{uu} = 0$. Let $D_i$ consists of all regions adjacent to region $i$. We assume the CAR model defined by the conditional probability density function

$$f(Z_i|Z_j, j \neq i) = \Big(\frac{d_i}{2\pi\delta_1}\Big)^{1/2} \exp\Big\{-\frac{d_i}{2\delta_1}\Big(Z_i - \rho_1 \bar{Z}_{-i}\Big)^2\Big\}, \tag{2}$$

where $\delta_1 > 0$, $|\rho_1| < 1$, $\bar{Z}_{-i} = d_i^{-1} \sum_{j \in D_i} Z_j$ and $d_i$ is the number of regions in $D_i$.

The limiting case where $\rho_1 = 1$ has been used by Besag *et al.* [13], Waller *et al.* [20], Ghosh *et al.* [16], among others. Sun, Tsutakawa and Speckman[24] have shown that when $\rho_1 = 1$, the joint distribution of $\mathbf{Z} = (Z_1, \cdots, Z_I)$ is not singular normal but partially informative and improper. For $|\rho_1| < 1$, they apply Besag's[3,p.201] result on an auto-regressive scheme to show that $\mathbf{Z}$ is multivariate normal with mean $\mathbf{0}$ and nonsingular covariance matrix $\delta_1(\mathbf{D} - \rho_1\mathbf{C})^{-1}$, where $\mathbf{D} = diag(d_1, \cdots, d_I)$.

We note that when $\rho_1 = 0$, the $Z_i's$ are independent and, when $0 \leq \rho_1 \leq 1$, $\rho_1$ may be interpreted as the shrinkage factor in the conditional mean,

$$E(Z_i|Z_j, j \neq i) = (1 - \rho_1) \times 0 + \rho_1 \bar{Z}_{-i},$$

which is the weighted average of the prior mean 0 and the average of the neighboring $Z_j$, with more (less) weight given to the prior mean when $\rho_1$ is small (large). Thus $\rho_1$ serves as an index of spatial dependence, a feature not present when $\rho_1$ is restricted to 1.

Similarly, we assume that $\mathbf{W} = (W_1, \cdots, W_I)'$ has a multivariate normal distribution with mean $\mathbf{0}$ and covariance matrix $\delta_2(\mathbf{D} - \rho_2\mathbf{C})^{-1}$ for $|\rho_2| < 1$. The conditional distribution of $W_i$ given $W_j, j \neq i$, is then normal with mean $\rho_2 \overline{W}_{-i}$ and variance $\delta_2/d_i$, where $\overline{W}_{-i} = d_i^{-1} \sum_{j \in D_i} W_j$. We note that correlation coefficients $\rho_1$ and $\rho_2$ are assumed to be independent of age group and time period.

For our illustration we have chosen to use the CAR model (2) for both $\mathbf{Z}$ and $\mathbf{W}$, which has, to the best our knowledge, not previously been used in disease mapping. This model not only has a proper joint distribution, but is simpler to implement in the MCMC algorithm than the model previously used by Besag *et al.*[13]. Moreover, unlike the previous model, one can assess the degree of correlations among neighboring regions through the posterior distribution of $\rho_1$ or $\rho_2$.

### 2.3 Choice of Priors

First define $\mathbf{p} = (p_{111}, \cdots, p_{11K}, p_{121}, \cdots, p_{IJK})'$, $\boldsymbol{\theta} = (\theta_1, \cdots, \theta_J)'$, $\boldsymbol{\mu} = (\mu_1, \cdots, \mu_J)'$, $\boldsymbol{\delta} = (\delta_1, \delta_2)'$ and $\boldsymbol{\rho} = (\rho_1, \rho_2)'$.

We assume that for given $(\boldsymbol{\theta}, \boldsymbol{\mu}, \mathbf{Z}, \mathbf{W})$, the first stage prior distributions of $p_{ijk}$ follows Equation (1), where $e_{ijk}$ are iid $N(0, \delta_0)$. The following conditional independence assumptions are needed.

- Given $(\boldsymbol{\theta}, \boldsymbol{\mu}, \mathbf{Z}, \mathbf{W})$, $\mathbf{p}$ is conditionally independent of $\boldsymbol{\delta}$ and $\boldsymbol{\rho}$;
- Given $\boldsymbol{\delta}$ and $\boldsymbol{\rho}$, $\mathbf{Z}$, $\mathbf{W}$ and $(\boldsymbol{\theta}, \boldsymbol{\mu}, \delta_0)$ are mutually independent;
- $\theta_1, \cdots, \theta_J, \mu_1, \cdots, \mu_J, \delta_1, \delta_2, \rho_1, \rho_2$ and $\delta_0$ are mutually independent.

To specify the Bayesian hierarchical model, we must specify the prior distributions of $\boldsymbol{\theta}$, $\boldsymbol{\mu}$, $\boldsymbol{\delta}$, $\boldsymbol{\rho}$, and $\delta_0$. In particular, the following conditional distributions are assumed.

- The age effect, $\theta_j \sim N(\xi_{mj}, \delta_{mj})$, $j = 1, \cdots, J$;
- The mean slope, $\mu_j \sim N(\xi_{sj}, \delta_{sj})$, $j = 1, \cdots, J$;
- $\delta_l \sim$ Inverse Gamma type $(a_l, b_l)$, $\quad l = 0, 1, 2$, with density

$$p(\delta_l) \propto \delta_l^{-(a_l+1)} e^{-b_l/\delta_l}, \quad -\infty < a_l < \infty, \text{ and } b_l \geq 0.$$

- $\rho_1$ and $\rho_2$ are iid uniform on the interval $(-1, 1)$.

Here the hyperparameters $(\xi_{mj}, \delta_{mj})$, $(\xi_{sj}, \delta_{sj})$, $(a_l, b_l)$ are fixed constants. When $a_l > 0$ and $b_l > 0$, $\delta_l$ has a proper distribution.

## 3 Propriety of Posterior Distribution

To complete the hierarchical model, we need to specify the hyperparameters $(\xi_{mj}, \delta_{mj})$, $(\xi_{sj}, \delta_{sj})$ and $(a_l, b_l)$. In practice, it may not be easy to find these hyperparameters. Alternatively, it is more convenient to find the noninformative priors for such a hierarchical model. The commonly used noninformative prior for $\boldsymbol{\theta}$ and $\boldsymbol{\mu}$ are flat priors, i.e. a constant prior for $\boldsymbol{\theta}$ and $\boldsymbol{\mu}$. Of course, this can be a limiting case when $\delta_{mj}$ and $\delta_{sj}$ go to $\infty$. Note that a flat prior for $\rho_1$ or $\rho_2$ was also assumed. Noninformative priors for $\delta_l$ are quite tricky. Following tradition, one may want to use the prior $1/\delta_l$ for $\delta_l$, i.e. $\log(\delta_l)$ has a constant prior. However, Sun, Tsutakawa and He[27] show that the posterior from using such a prior will be improper. For a linear mixed model we prefer using the prior for $\delta_0$ whose density is proportional to $1/\delta_0$ and constants priors for other variance components (cf. Sun, Tsutakawa and He[27]). However, in doing so in our model the propriety of the posterior may no longer hold. Our procedure for estimation is based on a theorem on the existence of the posterior, which is given in the appendix.

# 4 Estimation via MCMC

The number of parameters $(\mathbf{p}, \boldsymbol{\theta}, \boldsymbol{\mu}, \mathbf{Z}, \mathbf{W}, \boldsymbol{\delta}, \boldsymbol{\rho}, \delta_0)$ is $IJK + 2J + 2I + 2 + 2 + 1$. In our case, because $(I, J, K) = (115, 4, 4)$, there are $2,083$ parameters. Clearly, Bayesian computation via numerical integration is not feasible. Instead we used Gibbs sampling for the computation. For this, we need to sample successively from full conditional distributions. In the next subsection, we first list these full conditional distributions. Some of them are standard distributions such as normal and inverse gamma, while many others require sampling from log-concave densities.

## 4.1 Available Conditional Distributions and Algorithm

We have found all the full conditional distributions for the above hierarchical structure. The proof of these results are straight forward and thus omitted.

(i) For given $(\boldsymbol{\theta}, \boldsymbol{\mu}, \mathbf{Z}, \mathbf{W}, \boldsymbol{\delta}, \boldsymbol{\rho}, \delta_0)$, $\log(p_{111}), \cdots, \log(p_{11k}), \cdots, \log(p_{IJK}))$ are independent. For any $(i, j, k)$, the conditional density of $v_{ijk} = \log(p_{ijk})$ given $(\boldsymbol{\theta}, \boldsymbol{\mu}, \mathbf{Z}, \mathbf{W}, \boldsymbol{\delta}, \boldsymbol{\rho}, \delta_0)$ is proportional to

$$\exp\Big\{y_{ijk} v_{ijk} - n_{ijk} e^{v_{ijk}} - \frac{1}{2\delta_0}(v_{ijk} - a_{ijk})^2\Big\}, \tag{3}$$

where

$$a_{ijk} = \theta_j + Z_i + (\mu_j + W_i)(t_k - \bar{t}). \tag{4}$$

(ii) For any $j = 1, \cdots, J$, the conditional distribution of $\theta_j$ given $(\mathbf{p}, \boldsymbol{\mu}, \mathbf{Z}, \mathbf{W}, \boldsymbol{\delta}, \boldsymbol{\rho}, \delta_0)$ is

$$N\Big(\frac{\frac{\xi_{mj}}{\delta_{mj}} + \frac{1}{\delta_0}\sum_{i,k}(v_{ijk} - Z_i)}{(IK/\delta_0) + (1/\delta_{mj})},\ \frac{1}{(IK/\delta_0) + (1/\delta_{mj})}\Big).$$

(iii) Define $c_i = \delta_0^{-1}\sum_{j,k}(v_{ijk} - \theta_j)$, $\mathbf{c} = (c_1, \cdots, c_I)'$, the $I \times I$ identity matrix by $\mathbf{I}$, and $\mathbf{A}_i = \mathbf{D} - \rho_i \mathbf{C}$, $i = 1, 2$. Then the conditional distribution of $\mathbf{Z}$ given $(\mathbf{p}, \boldsymbol{\theta}, \boldsymbol{\mu}, \mathbf{W}, \boldsymbol{\delta}, \boldsymbol{\rho}, \delta_0)$ is

$$\mathrm{MVN_I}\Big(\Big(\frac{JK}{\delta_0}\mathbf{I} + \frac{1}{\delta_1}\mathbf{A}_1\Big)^{-1}\mathbf{c},\ \Big(\frac{JK}{\delta_0}\mathbf{I} + \frac{1}{\delta_1}\mathbf{A}_1\Big)^{-1}\Big). \tag{5}$$

(iv) The conditional distribution of $\delta_1$ given $(\mathbf{p}, \boldsymbol{\theta}, \boldsymbol{\mu}, \mathbf{Z}, \mathbf{W}, \delta_2, \boldsymbol{\rho}, \delta_0)$ is others is Inverse Gamma $(a_1 + \frac{1}{2}I,\ b_1 + \frac{1}{2}\mathbf{Z}'\mathbf{A}_1\mathbf{Z})$.

(v) The conditional density of $\rho_1$ given $(\mathbf{p}, \theta, \boldsymbol{\mu}, \mathbf{Z}, \mathbf{W}, \boldsymbol{\delta}, \rho_2, \delta_0)$ is proportional to

$$|\mathbf{D} - \rho_1 \mathbf{C}|^{1/2} \exp\Big\{\frac{\rho_1}{2\delta_1}\mathbf{Z}'\mathbf{C}\mathbf{Z}\Big\}. \tag{6}$$

(vi) Let $s_t^2 = \sum_{k=1}^K (t_k - \bar{t})^2$. Then the conditional distribution of $\mu_j$ given $(\mathbf{p}, \theta, \mathbf{Z}, \mathbf{W}, \boldsymbol{\delta}, \rho_2, \delta_0)$ is proportional to

$$N\Big(\frac{\frac{\xi_{sj}}{\delta_{sj}} + \frac{1}{\delta_0}[\sum_{i,k} v_{ijk}(t_k - \bar{t}) - s_t^2 \sum_i W_i]}{(Is_t^2/\delta_0) + (1/\delta_{sj})}, \ \frac{1}{(Is_t^2/\delta_0) + (1/\delta_{sj})}\Big).$$

(vii) Define $\eta_i = \delta_0^{-1}[\sum_{j,k} v_{ijk}(t_k - \bar{t}) - s_t^2 \sum_j \mu_j]$ and $\boldsymbol{\eta} = (\eta_1, \cdots, \eta_I)'$. Then the conditional distribution of $\mathbf{W}$ given $(\mathbf{p}, \theta, \mathbf{u}, \mathbf{Z}, \boldsymbol{\delta}, \rho_2, \delta_0)$ is

$$\mathrm{MVN_I}\Big(\Big(\frac{Js_t^2}{\delta_0}\mathbf{I} + \frac{1}{\delta_2}\mathbf{A}_2\Big)^{-1}\boldsymbol{\eta}, \ \Big(\frac{Js_t^2}{\delta_0}\mathbf{I} + \frac{1}{\delta_2}\mathbf{A}_2\Big)^{-1}\Big).$$

(viii) The conditional distribution of $\delta_2$ given $(\mathbf{p}, \theta, \boldsymbol{\mu}, \mathbf{Z}, \mathbf{W}, \delta_1, \boldsymbol{\rho}, \delta_0)$ is Inverse Gamma $(a_2 + \frac{1}{2}I, \ b_2 + \frac{1}{2}\mathbf{W}'\mathbf{A}_2\mathbf{W})$.

(ix) The conditional density of $\rho_2$ given $(\mathbf{p}, \theta, \boldsymbol{\mu}, \mathbf{Z}, \mathbf{W}, \boldsymbol{\delta}, \rho_1, \delta_0)$ is proportional to

$$|\mathbf{D} - \rho_2\mathbf{C}|^{1/2} \exp\Big\{\frac{\rho_2}{2\delta_2}\mathbf{W}'\mathbf{C}\mathbf{W}\Big\}.$$

(x) The conditional distribution of $\delta_0$ given $(\mathbf{p}, \theta, \boldsymbol{\mu}, \mathbf{Z}, \mathbf{W}, \boldsymbol{\delta}, \boldsymbol{\rho})$ is Inverse Gamma $(a_0 + \frac{1}{2}IJK, \ b_0 + \frac{1}{2}\sum_{i=1}^I \sum_{j=1}^J \sum_{k=1}^K (v_{ijk} - a_{ijk})^2)$, where $a_{ijk}$ is given by (4).

It is easy to generate samples from the normal and inverse gamma distributions given in (ii)–(iv), (vi)–(viii) and (x). We can show that if we use the parameter $p_{ijk}$ in the computation, the conditional density of $p_{ijk}$ given others is not log-concave. However, as shown below, $v_{ijk} = \log(p_{ijk})$, which is a monotone transformation of $p_{ijk}$, has a log-concave conditional density. Alternatively, we can simulate in terms of $e_{ijk}$ in place of $p_{ijk}$ or $v_{ijk}$. Using $v_{ijk}$ makes it easier to relate to the classical linear mixed model. We will also see that the conditional densities of $\rho_1$ and $\rho_2$ are log-concave. This makes it possible for us to use the adaptive rejection sampling method by Gilks and Wild[28] to sample from these log-concave densities.

**Lemma 1** *The conditional density of $v_{ijk} = \log(p_{ijk})$ given in (i) is log-concave.*

**Proof.** Let $h(v_{ijk})$ be the exponent in (3). The second derivative of $h$ with respect to $v_{ijk}$ is $-n_{ijk}\exp(v_{ijk}) - \delta_0^{-1}$, which is negative. □

We note that there are two ways to simulate the conditional distribution of $\mathbf{Z}$ (or $\mathbf{W}$). The conventional method is to perform this one $Z_i$ at a time. The other is to perform this on the joint distribution of $\mathbf{Z}$ as a block. From (iii) and (vii), we see that the mean and variance matrix of

conditional density of $\mathbf{Z}$ (or $\mathbf{W}$) involve the inverse of an $I \times I$ matrix, such as $\frac{JK}{\delta_0}\mathbf{I} + \frac{1}{\delta_1}\mathbf{A}_1$ (or $\frac{Js_t^2}{\delta_0}\mathbf{I} + \frac{1}{\delta_2}\mathbf{A}_2$). In our case, this requires computing inverses of $J+1$ such matrices within each Gibbs cycle. Moreover, matrix multiplications are also needed to simulate a sample of $(\delta_1, \rho_1)$ or $(\delta_2, \rho_2)$. To get accurate Bayesian estimators, a large number of cycles is needed. It would be nice to reduce the number of matrix multiplications and have a better computational formula which does not require finding inverses of such matrices. For our particular problem, it is simpler, for this reason, to sample the components of $\mathbf{Z}$ one at a time. For this we use the conditional distribution of $Z_i$ given $(Z_j, j \neq i, \mathbf{p}, \theta, \boldsymbol{\mu}, \mathbf{W}, \boldsymbol{\delta}, \boldsymbol{\rho}, \delta_0)$ by

$$N\Big(\frac{\frac{1}{\delta_0}\sum_{j,k}(v_{ijk} - \theta_j) + \frac{\rho_1}{\delta_1}\sum_l c_{il} Z_l}{(JK/\delta_0) + (d_i/\delta_1)},\ \frac{1}{(JK/\delta_0) + (d_i/\delta_1)}\Big).$$

Similarly, the conditional distribution of $W_i$ given $(W_j, j \neq i, \mathbf{p}, \theta, \boldsymbol{\mu}, \mathbf{Z}, \boldsymbol{\delta}, \boldsymbol{\rho}, \delta_0)$ is

$$N\Big(\frac{\frac{1}{\delta_0}\{\sum_{j,k} v_{ijk}(t_k - \bar{t}) - s_t^2 \sum_j \mu_j\} + \frac{\rho_2}{\delta_2}\sum_l c_{il} W_l}{(Js_t^2/\delta_0) + (d_i/\delta_2)},\ \frac{1}{(Js_t^2/\delta_0) + (d_i/\delta_2)}\Big).$$

**Lemma 2** *(a) Let $\lambda_1^*, \ldots, \lambda_I^*$ be eigenvalues of $\boldsymbol{CD}^{-1}$. The conditional density of $\rho_1$, given in (6), is equivalent to*

$$h(\rho_1) \propto \exp\Big\{\frac{1}{2}\sum_{i=1}^{I}\log(1 - \rho_1\lambda_i^*) + \frac{\rho_1}{2\delta_1}\boldsymbol{Z}'\boldsymbol{CZ}\Big\}, \quad \rho_1 \in (-1, 1). \tag{7}$$

*(b) The conditional density of $\rho_1$ in (6) is log-concave.*

**Proof.** Part (a) follows from the fact that $|\mathbf{D} - \rho_1\mathbf{C}| = |\mathbf{D}|\prod_{i=1}^{I}(1 - \rho_1\lambda_i^*)$. For part (b), the second derivative of $h(\rho_1)$ is

$$\frac{\partial^2}{\partial\rho_1^2}\log[h(\rho_1)] = -\frac{1}{2}\sum_{i=1}^{I}\frac{\lambda_i^{*2}}{(1 - \rho_1\lambda_i^*)^2} < 0.$$

This proves the results. □

Similarly, we can show that the conditional density of $\rho_2$ is also log-concave. Note that we only need to compute the eigenvalues and eigenvectors of $\mathbf{CD}^{-1}$ once at the beginning of the computation.

## 4.2 Numerical Results

### 4.2.1 Specification of Prior Distributions

In our numerical example, we have used two sets of prior distributions which are specified in Table 1. The first set is used in our main analysis of mortality rates and consists of relatively

diffuse but proper prior distributions. The second set is used to examine the robustness of the posterior distribution relative to changes in the prior and consists of noninformative distributions. A summary of the marginal posterior distributions of the parameters $(\boldsymbol{\theta}, \boldsymbol{\mu}, \boldsymbol{\delta}, \boldsymbol{\rho}, \delta_0)$ under the informative prior is presented in Table 2. The robustness issue will be addressed in Section 4.3.

#### 4.2.2 Age effects $\boldsymbol{\theta}$ and $\boldsymbol{\mu}$

The posterior means of $\theta_j$ indicate a rapid increase with respect to age, except the two oldest groups which have very similar values. The posterior means of $\mu_j$ show an increase with respect to age. The negative value for the youngest group indicates an overall decrease in mortality over the 20 year period. A more detailed pattern of the increases is seen in the posterior distributions of $\theta_j$ and $\mu_j$ plotted in Figure 2. One implication of this is that there is a positive interaction between age and time.

#### 4.2.3 Variance Components $\boldsymbol{\delta}$ and $\delta_0$

The variability in $\mathbf{Z}, \mathbf{W}$ and $\mathbf{e}$ may be seen in their variances, $\delta_1, \delta_2$, and $\delta_0$, whose posterior distributions are summarized in Table 2 and plotted in Figure 3. The concentration of the posterior distribution away from 0 for $\delta_0$ and $\delta_1$ indicates the importance of $e_{ijk}$ and $Z_i$ in the model. In particular, a positive $\delta_0$ indicates the lack of fit of the linear term in (1) without the extra variation $e_{ijk}$. The small values of $\delta_2$ indicates there is little variability in the changes over time among most counties. However, we will see some noteworthy exceptions in the disease maps of $W_i$ in Section 5. We have computed the ordinary sum squares of errors of the fitted loglinear model. About 64.68% of the total variation in $\log(\hat{p}_{ijk})$ is explained by the linear fit. That is to say, about 35.32% of total variation remains in the extra variation $\hat{e}_{ijk}$, which could be due to other effects that are not included in linear terms of the model.

#### 4.2.4 Spatial Correlations $\rho_i$

Table 2 and the plots of the posterior distributions of the $\rho's$ shown in Figure 4 indicate the spatial correlation $\rho_1$ for $\mathbf{Z}$ is clearly positive but $\rho_2$ for $\mathbf{W}$ has distribution widely spread about the origin.

The effect of $\rho_1$ and $\rho_2$ on $\mathbf{Z}$ may be seen by comparing the posterior means of $\mathbf{Z}$ with and without these parameters. Figure 5 gives scatter plots which suggest that the effect of $\rho_1$ on $\mathbf{Z}$ is present, but the effect of $\rho_2$ on $\mathbf{Z}$ is negligible.

The effect of $\rho_1$ and $\rho_2$ on $\mathbf{W}$ may be seen, similarly, by comparing the posterior means of $W_i$ with and without these parameters. The scatter plots in Figure 5 suggest the effect of $\rho_2$ on $\mathbf{W}$ but not on $\mathbf{Z}$.

Finally, the effect of $\rho_1$ and $\rho_2$ on $p_{ijk}$ may be seen by comparing $\hat{p}_{ijk}$ and $\hat{p}^*_{ijk}$, the posterior means with and without $(\rho_1, \rho_2)$, respectively. The scatter plots of these means, shown in Figure 6 have a correlation of 0.9989 and show negligible differences, except for in a few isolated cases. The absolute relative error $|\hat{p}_{ijk} - \hat{p}^*_{ijk}|/\hat{p}_{ijk}$ is less than 5% in 93% of the cases. Moreover, the correlations for the 16 age and time groups range from 0.981 to 0.988.

One conclusion which we have drawn from our graphical analysis is that, spatial correlations among counties exist in terms of $Z_i$, but not in terms of the change over time $W_i$. The conclusion is supported by the Bayes factors which were computed using the method of Meng and Wong (1996). For the three reduced models:

$$\text{(i) } \rho_1 = 0, \qquad \text{(ii) } \rho_2 = 0, \qquad \text{and (iii) } \rho_1 = \rho_2 = 0,$$

the values of the Bayes factors (in support of the full model) are $31,912$; $2.1$; and $62,028$, respectively. Although there are strong correlations among the $Z_i$, the effect of these correlations on $p_{ijk}$ is relatively small due to the significant effect of $e_{ijk}$.

We have considered using the Bayes factor for fitting other models by employing several methods, including those proposed by Chib[30] and Newton and Raftery[31]. However, we encountered difficulties due to the instability of the approximations arising from an occasional extreme value in the simulation.

### 4.3 Robustness in terms of Choice of Priors

We examine the sensitivity of the posterior distribution relative to the choice of prior distributions by comparing the results of the informative prior with the noninformative prior specified in Table 1. We note that the prior distributions for $\delta_0$, $\delta_1$ and $\delta_2$ in the noninformative case are improper but satisfy the conditions of the theorem.

A simple comparison of the results based on the two priors can be made by comparing the summary statistics in Table 2 with those in Table 3. We note that there is general agreement in the marginal posterior distributions of $(\boldsymbol{\theta}, \boldsymbol{\mu}, \boldsymbol{\delta}, \boldsymbol{\rho}, \delta_0)$. Moreover, the posterior means of $p_{ijk}$ under the informative and noninformative priors, $\hat{p}^*_{ijk}$ and $\hat{p}_{ijk}$, are quite close, with absolute relative error always less than 0.006 and with 95.1% less than 0.003 (see Figure 6). In using "more" informative

priors for $\delta_0$, there were some changes in the posterior of $\delta_l$ and $\rho_l$, but negligible changes in the posteriors of the fixed parameters, $\boldsymbol{\theta}$ and $\boldsymbol{\mu}$. We therefore feel that our Bayesian results for the present example are quite robust with respect to the choice of the hyperparameters, provided they represent relatively diffuse prior distributions.

### 4.4 Convergence Diagnostic

In monitoring the convergence of Gibbs sampling, we have run three parallel MCMC chains with widely different starting values and used graphical monitoring of the chains for a representative subset of the parameters including large, moderate and small mortality rates ($p_{95,3,1}, p_{79,2,4}, p_{113,4,3}$), and other parameters such as $\theta_2, \mu_1, Z_{79}, W_{113}, \delta_2, \rho_2$, and $\delta_0$. Figure 7 shows the updated means of Gibbs sampling, with respect to the number of iterations, for the subset of parameters. It appears that the estimators are quite stable after about $20,000$ iterations.

Gelman and Rubin[25] have suggested monitoring the estimated scale reduction factor $\hat{R}$ of the parallel sequences and to choose the Gibbs cycles so that $\hat{R}$ is less than 1.2. In our cases, the estimated scale reduction factor $\hat{R}$ are less then 1.05 when the Gibbs sample sizes are larger than $20,000$. The same conclusion was shown from Figure 7. Due to the large number of parameters for our estimates, we used a sample of $40,000$, after discarding the burn-in sample of $10,000$. For illustration, we used all $50,000$ for Figure 7.

## 5 Data Analysis via Disease Mapping

In this section we present several mappings to illustrate the use of different components of the hierarchical model and discuss specific topics listed in Section 1. We first present the map of the estimated annual mortality rates per $10^5$ population to illustrate the overall smoothing attained by the hierarchical model, the patterns of possible changes over time, and the concentration of regions of high and low rates. We then discuss the map of the regional effects, changes over time by age, and then the extra variation.

Using the full model (1), Figures 8(a) and 8(b) map the posterior means and standard deviations of the rate $p_{ijk}$, which have been rescaled to represent the annual rates per $10^5$ population. We first note that most of the extreme rates noted in Figure 1 have disappeared. In particular the isolated cases of high rates among the two lower age groups have disappeared and the isolated lower rates among the two high age groups have also disappeared. The trend in increasing rates over age can

now be seen clearly. Moreover we clearly notice how the rates have increased over time for the two older groups. The decrease over time for the youngest group discussed earlier is less clear due to the small range of values. There are a few counties where the trends are in the opposite direction. We also notice geographical clusters of similar rates such as the lower rates in the north and higher rates in the southeast. The overall pattern we notice in Figure 8 (a) is the general increase in rate from the upper left to the lower right of the figure. This is consistent with the simultaneous increase in $\theta_j$ and $\mu_j$ observed in Figure 2.

Figure 9 (a) shows a map of the regional effect $Z_i$ in terms of $e^{Z_i}$, which corresponds to the spatially correlated component of the relative risks and should be close to 1 for the "average" county. We can now notice clusters with lower levels appearing in the north. Higher levels are seen among the adjacent counties in the southeast quadrant. The highest rate occurs in St. Louis city, located in the middle of the eastern boundary of the state, and there are few other isolated counties with higher risks.

Figure 9 (b) is a map of $W_i$, the rate of change over time, without the age effect. We note clusters of higher $W_i$ in the southwest and south central regions. St. Loius City which had the largest $Z_i$ now shows a closer to average value of $W_i$, Osage County, located near the center of the state, has conspicuously low $Z_i$ and average $W_i$ values, unlike its neighbors.

In our model, the change over time has two linear components, one due to age ($\mu_j$) and the other space ($W_i$). The age-time interaction is quite evident with $\mu_j$ increasing with age. The space-time interaction is present but less evident due to the concentration of $W_i$ about zero. We note that this clustering is close in part to the "shrinkage" property of estimates commonly associated with Bayesian hierarchical models, which is enhanced in our case by the large number of counties with small populations.

Figure 10 shows maps of the extra variation, $e_{ijk}$, which remains after the model is fit using the linear terms. There could be a number of reasons why the linear model (1) with $e_{ijk} \equiv 0$ is inadequate. One is the nonlinearity of the temporal effect on the log rate. Another is the possible interaction between age and county. There could be a number of other reasons such as race, environment, and regional variations in smoking habits, for which we have no available data. We note that other than a few isolated spots the extra variations are fairly evenly dispersed over time and space. One interesting case is the oldest group in St. Louis county (adjacent to St. Louis city) where the $e_{ijk}$ is high for the two middle periods and closer to 0 for the first and fourth periods. The inclusion of $e_{ijk}$ allows for such lack of linearity which cannot be individually addressed with

the current data alone. Environmental effects on lung cancer are often local, both spatially and temporally, and affect different age group differently. We feel that a careful examination of $e_{ijk}$ can be very helpful in a study of such local changes. Obviously these mappings suggest regions with potential problems and a more extensive and thorough analysis is required before health related conclusions can be drawn.

## Appendix. Theorem 1 and a Corollary

We express our model in terms of a more general notation and give sufficient conditions, including those on the noninformative prior, which imply that the joint posterior is proper.

Let $N = IJK$. Relabel $\mathbf{Y} = (Y_{1,1,1}, \cdots, Y_{1,1,K}, Y_{1,2,1}, \cdots, Y_{I,J,K})$ as $(Y_1, \cdots, Y_N)$, $\mathbf{V} = (\log(p_{1,1,1}),$ $\cdots, \log(p_{1,1,K}), \log(p_{1,2,1}), \cdots, \log(p_{I,J,K}))$ as $(V_1, \cdots, V_N)$. Let $\mathbf{1}_I$ be the $I \times 1$ vector, whose components are all ones. Define two design matrices by

$$\begin{aligned} \mathbf{X}_1 &\equiv \Big(\mathbf{1}_I \otimes \mathbf{I}_J \otimes \mathbf{1}_K,\ \mathbf{1}_I \otimes \mathbf{I}_J \otimes (\mathbf{t} - \bar{t}\mathbf{1}_K)\Big) = (x_{1,1}, \cdots, x_{1,N})'; \\ \mathbf{X}_2 &\equiv \Big(\mathbf{I}_I \otimes \mathbf{1}_J \otimes \mathbf{1}_K,\ \mathbf{I}_I \otimes \mathbf{I}_J \otimes (\mathbf{t} - \bar{t}\mathbf{1}_K)\Big) = (x_{2,1}, \cdots, x_{2,N})', \end{aligned}$$

whose dimensions are $N \times 2J$ and $N \times I(J+1)$, respectively. Here $\mathbf{A} \otimes \mathbf{B}$ is the Kronecker or direct product of two matrices $\mathbf{A}$ and $\mathbf{B}$. The model (1) is equivalent to

$$V_i = x'_{1,i}(\boldsymbol{\theta}', \boldsymbol{\mu}')' + x'_{2,i}(\mathbf{Z}', \mathbf{W}')' + E_i,$$

where, $(E_1, \cdots, E_N) = (e_{1,1,1}, \cdots, e_{1,1,K}, e_{1,2,1}, \cdots, e_{I,J,K})$. This is equivalent to the matrix equation,

$$\mathbf{V} = \mathbf{X}_1 \begin{pmatrix} \boldsymbol{\theta} \\ \boldsymbol{\mu} \end{pmatrix} + \mathbf{X}_2 \begin{pmatrix} \mathbf{Z} \\ \mathbf{W} \end{pmatrix} + \mathbf{e}.$$

**Theorem 1** *Assume that the following conditions hold.*

*(a) The priors for* $\boldsymbol{\theta}$ *and* $\boldsymbol{\mu}$ *are proportional to constants.*

*(b) The data* $Y_i$ *and the hyperparameters* $(a_l, b_l), l = 0, 1, 2$ *satisfy the following conditions.*

*(b1) Either* $b_j > 0$ *or* $a_j < b_j = 0$, *j=1,2;*

*(b2)* $I + a_j > 2$, $j = 1, 2$;

*(b3)* $b_0 > 0$;

*(b4) There exists* $Y_{l_1}, \cdots, Y_{l_n}$ $(1 \le l_1 < \cdots < l_n \le N;\ 2J \le n \le N)$ *such that*

*(b4.1)* $Y_{l_j} > 0$;

*(b4.2) The* $n \times 2J$ *matrix* $\boldsymbol{X}_1^* = (x_{1,l_1}, \cdots, x_{1,l_n})'$ *is of full rank* $(2J)$*;*

*(b4.3)* $\frac{1}{2}n - J + a_0 + \min(0, a_1) + \min(0, a_2) > 0$.

*Then the joint posterior distribution of* $(\boldsymbol{p}, \boldsymbol{\theta}, \boldsymbol{\mu}, \boldsymbol{Z}, \boldsymbol{W}, \boldsymbol{\delta}, \boldsymbol{\rho}, \delta_0)$ *is proper.*

The proof of the Theorem can be found in the Appendix. Some examples of improper priors for $\delta_l$, $l = 1, 2$, satisfying the conditions are

(i) $p(\delta_l) \propto \delta_l^{-1/2}$, when $(a_l, b_l) = (-\frac{1}{2}, 0)$;

(ii) $p(\delta_l) \propto 1$, when $(a_l, b_l) = (-1, 0)$.

The following immediate corollary will be used in our illustration.

**Corollary 1** *Under the assumptions of (a) and (b4) in Theorem 1, if* $(a_l, b_l), l = 0, 1, 2$, *are all positive constants, then the joint posterior distribution of* $(\boldsymbol{p}, \boldsymbol{\theta}, \boldsymbol{\mu}, \boldsymbol{Z}, \boldsymbol{W}, \boldsymbol{\delta}, \boldsymbol{\rho}, \delta_0)$ *is proper.*

This is the case where all prior distributions are proper, except for these for $\boldsymbol{\theta}$ and $\boldsymbol{\mu}$.

Note that $Y_i$ given $V_i$ has Poisson distribution with mean $m_i e^{V_i}$, where $(m_1, \cdots, m_N) = (n_{1,1,1}, \cdots, n_{1,1,K}, \cdots, n_{I,J,K})$ and $V_i = \log(p_i)$. Let $f_i(Y_i|V_i) = \exp(V_i Y_i - m_i e^{V_i})$ be the likelihood function of $V_i$. Without loss of generality, assume that $l_j = j$, $j = 1, \cdots, n$. It is easily shown that

$$\begin{cases} \int f_j(Y_j|V_j) dV_j < \infty, & j = 1, \cdots, n; \\ f_j(Y_j|V_j) < M, & j = n+1, \cdots, N, \end{cases} \tag{8}$$

for some constant $M$. Let $V^* = (V_1, \cdots, V_n)$. The posterior density of $(\mathbf{V}, \boldsymbol{\theta}, \boldsymbol{\mu}, \mathbf{Z}, \mathbf{W}, \boldsymbol{\delta}, \boldsymbol{\rho}, \delta_0)$ given $\mathbf{Y}$ is

$$\begin{aligned} & p(\mathbf{V}, \boldsymbol{\theta}, \boldsymbol{\mu}, \mathbf{Z}, \mathbf{W}, \boldsymbol{\delta}, \boldsymbol{\rho}, \delta_0) \\ \propto & \prod_{i=1}^{n} f_i(Y_i|V_i) \prod_{i=n+1}^{N} \left( \delta_0^{-\frac{1}{2}} \exp\left\{ -\frac{1}{2\delta_0} [V_i - x_{1i}'(\boldsymbol{\theta}', \boldsymbol{\mu}')' - x_{2i}'(\mathbf{Z}', \mathbf{W}')']^2 \right\} \right) G^*. \end{aligned}$$

where $G^*$ is the joint prior density of $(\mathbf{V}^*, \boldsymbol{\theta}, \boldsymbol{\mu}, \mathbf{Z}, \mathbf{W}, \boldsymbol{\delta}, \boldsymbol{\rho}, \delta_0)$. Integrating with respect to $(V_{n+1}, \cdots, V_N)$, it follows from (8) that

$$p(\mathbf{V}^*, \boldsymbol{\theta}, \boldsymbol{\mu}, \mathbf{Z}, \mathbf{W}, \boldsymbol{\delta}, \boldsymbol{\rho}, \delta_0) \propto G^*.$$

Define the design matrices based on the first $n$ observations by $\mathbf{X}_1^* = (x_{1,1}, \cdots, x_{1,n})'$, $\mathbf{X}_2^* = (x_{2,1}, \cdots, x_{2,n})'$, and $\mathbf{X}^* = (\mathbf{X}_1^*, \mathbf{X}_2^*)$. Let $(\mathbf{X}^{*\prime}\mathbf{X}^*)^-$ denote the generalized inverse of the matrix

$(\mathbf{X}'\mathbf{X})$. It follows from Theorem 2 of Sun, Tsutakawa and He[27] that

$$\begin{aligned}\int G^* d\boldsymbol{\theta} d\boldsymbol{\mu} d\boldsymbol{\delta} d\boldsymbol{\rho} &\leq H\delta_0^{-\frac{1}{2}(n-2J)-a_+-1}\exp\Big\{-\frac{1}{2\delta_0}\mathbf{V}'[\mathbf{I}_n-\mathbf{X}^*(\mathbf{X}^{*\prime}\mathbf{X}^*)^-\mathbf{X}^{*\prime}]\mathbf{V}-\frac{b_0}{\delta_0}\Big\}\\ &\leq H\delta_0^{-\frac{1}{2}(n-2J)-a_+-1}\exp\Big\{-\frac{b_0}{\delta_0}\Big\}, \qquad (9)\end{aligned}$$

where $a_+ = a_0 + a_1 + a_2$, and $H$ is a constant depending only on the adjacency matrix $\mathbf{C}$ and the design matrix $\mathbf{X}^*$. The last inequality holds because $\mathbf{I}_n - \mathbf{X}^*(\mathbf{X}^{*\prime}\mathbf{X}^*)^-\mathbf{X}^{*\prime}$ is a nonnegative definite matrix. From the assumptions (b3) and (b4.3), the right hand side of (9) is integrable with respect to $\delta_0$. The result follows immediately.

## Acknowledgments

The research was supported in part by the US Federal Highway Administration, under subcontract E77690017-3Y from Los Alamos National Laboratory to the National Institute of Statistical Sciences. This work was also partially supported by a Research Board award from the University of Missouri-System. The work by He was partially supported by Federal Aid in Wildlife Restoration Project W-13-R. The authors wish to thank two referees for several helpful comments.

## References

1. Gelfand, A.E. and Smith, A.F.M. 'Sampling based approaches to calculating marginal densities', *Journal of the American Statistical Association*, **85**, 398-409 (1990).
2. Tanner, M.A. *Tools for Statistical Inference: Observed Data and Data Augmentation Methods.* Lecture Notes in Statistics, **67**, Springer Verlag, NY (1993).
3. Breslow, N.E. and Day, N.E. 'Indirect standardization and multiplicative models for rates, with reference to the age adjustment of cancer incidence and relative frequency data', *Journal of Chronic Diseases*, **28**, 289-303 (1975).
4. Manton, K., Woodbury, G., and Stallard, M.A. 'A variance components approach to categorical data models with heterogeneous cell populations: Analysis of spatial gradients in lung cancer mortality rates in North Carolina counties', *Biometrics*, **37**, 259-269 (1981).
5. Tsutakawa, R. K., Shoop, G. L. and Marienfeld, C. 'Empirical Bayes estimation of cancer mortality rates', *Statist. in Med.*, **4**, 201-212 (1985).
6. Tsutakawa, R.K. 'Mixed model for analysing geographic variability in mortality rates', *Journal of the American Statistical Association*, **83**, 117-130 (1988).
7. Whittle, P. 'On stationary process in the plane', *Biometrika*, **41**, 434-449 (1954).
8. Besag, J. 'Spatial interaction and the statistical analysis of lattice systems (with discussion)', *J. Roy. Statist. Soc. Ser. B*, **36**, 192-236 (1974).
9. Clayton, D. and Kaldor, J. 'Empirical Bayes estimates of age-standardized relative risks for use in disease mapping', *Biometrics*, **43**, 671-681 (1987).
10. Cliff, D. and Ord, J.K. *Spatial processes: models and applications.* Pion, London (1981).

11. Cressie, N. A.C. *Statistics for spatial data.* John Wiley & Sons, N.Y. (1993).
12. Clayton, D. and Bernardinelli, L. 'Bayesian methods for disease mapping disease risks', in Geographical and Environmental Epidemiology, editors, P. Elliot *et al.*, Oxford University Press, New York (1992).
13. Besag, J., York, J. and Mollié, A. 'Bayesian image restoration, with two applications in spatial statistics (with discussion)', *Ann. Inst. Statist. Math.*, **43**, 1-59 (1991).
14. Marshall, R.J. 'A review of methods for the statistical analysis of spatial patterns of disease', *J. Roy. Statist. Soc. Ser. A*, **154**, 421-441 (1991).
15. Carlin, B.P. and Louis, T.A. *Bayes and Empirical Bayes Methods for Data Analysis,* Chapman and Hall, London (1996).
16. Ghosh, M., Natarajan, K., Stroud, T.W.F., and Carlin, B.P. 'Generalized linear models for small area estimation', *Journal of the American Statistical Association*, **93**, 273-282 (1998).
17. Lee, W. and Lin, R.S. 'Autoregressive age-period-cohort models', *Statistics in Medicine*, **15**, 273-281 (1996).
18. Bernardinelli, L., Clayton, D. and Montomoli, C. 'Bayesian estimates of disease maps: how important are priors?' *Statistics in Medicine*, **14**, 2411-2431 (1995).
19. Bernardinelli, L., Clayton, D., Pascutto, C., Montomoli, C., Ghislandi, M., and Songini, M. 'Bayesian analysis of space-time variation in disease risk', *Statistics in Medicine*, **14**, 2433-2443 (1995).
20. Waller, L.A., Carlin, B.P., Xia, H., and Gelfand, A.E. 'Hierarchical spatio-temporal mapping of disease rates', *Journal of the American Statistical Association*, **92**, 607-617 (1997).
21. Besag, J., Green, P., Higdon, D., and Mengerson, K. 'Bayesian computation and stochastic systems', *Statistical Science,*, **10**, 1-41 (1995).
22. Xia, H. and Carlin, B.P. 'Spatial-temporal models with errors in covariates: mapping Ohio lung cancer mortality', *Statist. in Med.*, **17**, 2025-2043 (1998).
23. Knorr-Held, L. and Besag, J. 'Modeling risk from a disease in time and space', *Statist. in Med.*, **17**, 2045-2060 (1998).
24. Sun, D., Tsutakawa, R.K. and Speckman, P.L. 'Bayesian inference for CAR (1) models with noninformative priors', *Biometrika,* No. 2, in press (1999).
25. Gelman, A. and Rubin, D.B. 'Inference from iterative simulation using multiple sequences (with discussion)', *Statistical Science*, **7**, 457-511 (1992).
26. Waller, L.A., Carlin, B.P., and Xia, H. 'Structuring correlation within hierarchical spatio-temporal models for disease rates', In *Modeling Longitudinal and Spatially Correlated Data: Lecture Notes in Statistics 122,* T.G. Gregore, D.R. Brillinger, P.J. Diggle, E. Russek-Cohen, W.G. Warren, R.D. Woldinger, eds. Springer Verlag, New York, p.309-319 (1997).
27. Sun, D., Tsutakawa, R.K. and He, Z. 'Propriety of posteriors with improper priors in hierarchical linear mixed models', Revised for *Statistica Sinica* (1998).
28. Gilks, W.R. and Wild, P. 'Adaptive rejection sampling for Gibbs sampling', *Applied Statistics*, **41,** 337-348 (1992).
29. Meng,X. and Wong,W. H. 'Simulating ratios of normalizing constants via a simple identity: a theoretical exploration', *Statistica Sinica,* **6**, 831-860 (1996).
30. Chib, S. 'Marginal likelihood from the Gibbs outputs.' *Journal of the American Statistical Association*, **90**, 1313-1321 (1995).
31. Newton, M.A. and Raftery, A.E. 'Approximate Bayesian inference by the weighted likelihood bootstrap (with discussion)', *J. Roy. Statist. Soc. B*, **56**, 3-48 (1994).

Table 1. Choice of Hyperparameters.

| Parameter | Hyperparameter | Informative | Noninformative |
|---|---|---|---|
| $\theta_1$ | $(\xi_{m1}, \delta_{m1})$ | $(-5.684, 0.2880)$ | $(0, \infty)$ |
| $\theta_2$ | $(\xi_{m2}, \delta_{m2})$ | $(-4.515, 0.2616)$ | $(0, \infty)$ |
| $\theta_3$ | $(\xi_{m3}, \delta_{m3})$ | $(-3.877, 0.2620)$ | $(0, \infty)$ |
| $\theta_4$ | $(\xi_{m4}, \delta_{m4})$ | $(-3.805, 0.2663)$ | $(0, \infty)$ |
| $\mu_1$ | $(\xi_{s1}, \delta_{s1})$ | $(-0.01226, 14.12 \times 10^{-4})$ | $(0, \infty)$ |
| $\mu_2$ | $(\xi_{s2}, \delta_{s2})$ | $(0.01276, 6.503 \times 10^{-4})$ | $(0, \infty)$ |
| $\mu_3$ | $(\xi_{s3}, \delta_{s3})$ | $(0.01821, 5.009 \times 10^{-4})$ | $(0, \infty)$ |
| $\mu_4$ | $(\xi_{s4}, \delta_{s4})$ | $(0.03245, 6.265 \times 10^{-4})$ | $(0, \infty)$ |
| $\delta_1$ | $(a_1, b_1)$ | $(2.30, 0.276)$ | $(-1, 0.0)$ |
| $\delta_2$ | $(a_2, b_2)$ | $(0.10, 10^{-5})$ | $(-1, 0.0)$ |
| $\delta_0$ | $(a_0, b_0)$ | $(4.25, 0.195)$ | $(0, 0.1)$ |
| $\rho_1$ | | uniform$(-1, 1)$ | uniform$(-1, 1)$ |
| $\rho_2$ | | uniform$(-1, 1)$ | uniform$(-1, 1)$ |

Table 2. Posterior quantities based on 50000 Gibbs cycles under the informative prior.

| | Min. | 1st Qua. | Median | Mean | 3rd Qua. | Max. | Std. Dev. |
|---|---|---|---|---|---|---|---|
| $\theta_1$ | $-5.785$ | $-5.719$ | $-5.688$ | $-5.6876$ | $-5.660$ | $-5.531$ | 0.03892 |
| $\theta_2$ | $-4.703$ | $-4.546$ | $-4.517$ | $-4.5167$ | $-4.484$ | $-4.375$ | 0.03496 |
| $\theta_3$ | $-4.066$ | $-3.907$ | $-3.879$ | $-3.8785$ | $-3.843$ | $-3.728$ | 0.03391 |
| $\theta_4$ | $-3.997$ | $-3.836$ | $-3.807$ | $-3.8068$ | $-3.773$ | $-3.656$ | 0.03503 |
| $\mu_1$ | $-0.02878$ | $-0.01483$ | $-0.01226$ | $-0.01223$ | $-0.00966$ | 0.00368 | 0.003751 |
| $\mu_2$ | $-0.00009$ | 0.01103 | 0.01274 | 0.01273 | 0.01441 | 0.02340 | 0.002545 |
| $\mu_3$ | $-0.00806$ | 0.01684 | 0.01832 | 0.01833 | 0.01985 | 0.02745 | 0.002538 |
| $\mu_4$ | 0.02181 | 0.03075 | 0.03245 | 0.03245 | 0.03416 | 0.04371 | 0.002561 |
| $\delta_1$ | 0.02351 | 0.04823 | 0.05537 | 0.05662 | 0.06375 | 0.12670 | 0.011742 |
| $\delta_2$ | $1.22e-6$ | $1.74e-5$ | $3.65e-5$ | $5.60e-5$ | $7.35e-5$ | 0.00065 | $5.787e-5$ |
| $\delta_0$ | 0.00773 | 0.01277 | 0.01404 | 0.01412 | 0.01537 | 0.02322 | 0.001907 |
| $\rho_1$ | 0.2202 | 0.8700 | 0.9236 | 0.90314 | 0.9590 | 0.9900 | 0.07714 |
| $\rho_2$ | $-0.9900$ | $-0.3859$ | 0.1657 | 0.11834 | 0.6579 | 0.9900 | 0.59080 |

Table 3. Posterior quantities based on 50000 Gibbs cycles under the noninformative prior.

| | Min. | 1st Qua. | Median | Mean | 3rd Qua. | Max. | Std. Dev. |
|---|---|---|---|---|---|---|---|
| $\theta_1$ | $-5.883$ | $-5.719$ | $-5.689$ | $-5.6882$ | $-5.657$ | $-5.567$ | 0.05119 |
| $\theta_2$ | $-4.717$ | $-4.547$ | $-4.518$ | $-4.5176$ | $-4.488$ | $-4.304$ | 0.04955 |
| $\theta_3$ | $-4.071$ | $-3.908$ | $-3.879$ | $-3.8792$ | $-3.850$ | $-3.665$ | 0.04932 |
| $\theta_4$ | $-4.014$ | $-3.837$ | $-3.807$ | $-3.8077$ | $-3.778$ | $-3.587$ | 0.04986 |
| $\mu_1$ | $-0.02601$ | $-0.01473$ | $-0.01213$ | $-0.01210$ | $-0.00949$ | 0.00339 | 0.003826 |
| $\mu_2$ | $-0.00141$ | 0.01099 | 0.01272 | 0.01272 | 0.01446 | 0.02381 | 0.002578 |
| $\mu_3$ | $-0.00887$ | 0.01676 | 0.01829 | 0.01830 | 0.01981 | 0.02883 | 0.002285 |
| $\mu_4$ | 0.02112 | 0.03073 | 0.03248 | 0.03246 | 0.03420 | 0.04383 | 0.002602 |
| $\delta_1$ | 0.02519 | 0.04818 | 0.05523 | 0.05645 | 0.06329 | 0.13540 | 0.011618 |
| $\delta_2$ | $1.26e-6$ | $1.76e-5$ | $3.76e-5$ | $5.73e-5$ | $7.76e-5$ | 0.00063 | $5.701e-5$ |
| $\delta_0$ | 0.00808 | 0.01283 | 0.01405 | 0.01415 | 0.01537 | 0.02236 | 0.001878 |
| $\rho_1$ | 0.0512 | 0.8717 | 0.9251 | 0.9251 | 0.9603 | 0.9900 | 0.07617 |
| $\rho_2$ | $-0.9900$ | $-0.4028$ | 0.1415 | 0.1025 | 0.6356 | 0.9899 | 0.58896 |