



---

Generalized Linear Models for Small-Area Estimation

Author(s): Malay Ghosh, Kannan Natarajan, T. W. F. Stroud and Bradley P. Carlin

Source: *Journal of the American Statistical Association*, Vol. 93, No. 441 (Mar., 1998), pp. 273-282

Published by: [American Statistical Association](#)

Stable URL: <http://www.jstor.org/stable/2669623>

Accessed: 14/06/2013 10:31

---

Your use of the JSTOR archive indicates your acceptance of the Terms & Conditions of Use, available at <http://www.jstor.org/page/info/about/policies/terms.jsp>

JSTOR is a not-for-profit service that helps scholars, researchers, and students discover, use, and build upon a wide range of content in a trusted digital archive. We use information technology and tools to increase productivity and facilitate new forms of scholarship. For more information about JSTOR, please contact support@jstor.org.



American Statistical Association is collaborating with JSTOR to digitize, preserve and extend access to *Journal of the American Statistical Association*.

<http://www.jstor.org>

# Generalized Linear Models for Small-Area Estimation

Malay GHOSH, Kannan NATARAJAN, T. W. F. STROUD, and Bradley P. CARLIN

Bayesian methods have been used quite extensively in recent years for solving small-area estimation problems. Particularly effective in this regard has been the hierarchical or empirical Bayes approach, which is especially suitable for a systematic connection of local areas through models. However, the development to date has mainly concentrated on continuous-valued variates. Often the survey data are discrete or categorical, so that hierarchical or empirical Bayes techniques designed for continuous variates are inappropriate. This article considers hierarchical Bayes generalized linear models for a unified analysis of both discrete and continuous data. A general theorem is provided that ensures the propriety of posteriors under diffuse priors. This result is then extended to the case of spatial generalized linear models. The hierarchical Bayes procedure is implemented via Markov chain Monte Carlo integration techniques. Two examples (one featuring spatial correlation structure) are given to illustrate the general method.

KEY WORDS: Hierarchical model; Markov chain Monte Carlo; Posterior propriety; Spatial statistics.

## 1. INTRODUCTION

Bayesian methods have been used quite extensively in recent years for solving small-area estimation problems. Particularly effective in this regard have been the hierarchical Bayes (HB) and empirical Bayes (EB) approaches, which are especially suitable for a systematic connection of local areas through the model. For the general theory as well as specific applications of the HB and EB methods for small-area estimation, relevant work includes that of Datta and Ghosh (1991), Fay and Herriot (1979), Ghosh and Lahiri (1987, 1992), Ghosh and Meeden (1986), Prasad and Rao (1990), and Stroud (1987, 1991), among others. Ghosh and Rao (1994) have provided a review of many of these results.

But development to date has concentrated mainly on continuous-valued variates. Often the survey data are discrete or categorical, for which the HB or EB analysis suitable for continuous variates is not appropriate. Recently, some work has begun to appear on the Bayesian analysis of binary survey data. Dempster and Tomberlin (1980), Farrell, MacGibbon, and Tomberlin (1997) and MacGibbon and Tomberlin (1989) have obtained small area estimates of proportions via EB techniques, whereas Malec, Sedransk, and Tompkins (1993) found the predictive distributions of a linear combination of binary random variables using a HB technique. Stroud (1991) also developed a general HB methodology for binary data, and Nandram and Sedransk (1993) suggested Bayesian predictive inference for binary data from a two-stage cluster sample. Subsequently,

Stroud (1994) provided a comprehensive treatment of binary survey data encompassing simple random, stratified, cluster, and two-stage sampling, as well as two-stage sampling within strata.

The binary models constitute a subclass of generalized linear models that are often used for a unified analysis of both discrete and continuous data. Section 2 presents a general account of how HB generalized linear models (GLMs) can be used for small-area estimation. The section begins with a general description of HB GLMs. Sufficient conditions are provided for the joint posterior distribution of the parameters of interest to be proper under the proposed hierarchical models. The Bayes procedure is implemented via Markov chain Monte Carlo (MCMC) integration techniques—in particular, using the Gibbs sampler. Next, this section contains a discussion of some general multi-category models that may be handled indirectly by methods of this section, even though in their natural multinomial formulation they do not fit into the univariate GLM framework. We also point out that in contrast to the work of Stroud (1994), who used the Brooks (1984) method for approximating numerical integrals, we use exact MCMC integration techniques. We conclude this section by considering some spatial GLMs and find sufficient conditions that ensure the propriety of the posterior. We also point out a common HB model for this situation that actually leads to an improper posterior.

Section 3 contains the analysis of two real datasets. The first consists of responses to the question “Have you experienced any negative impact of exposure to health hazards in the workplace?” based on a 1991 sample of all persons in 15 geographic regions of Canada (Statistics Canada 1992). For each region, workers were classified by age ( $\leq 40$  or  $> 40$ ) and sex (male or female). The responses were classified into four categories: (1) yes, (2) no, (3) not exposed, and (4) not applicable or not stated. The objective is to estimate the proportion of workers in each of the four categories for every one of the  $15 \times 2 \times 2 = 60$  groups cross-classified by 15 geographic regions and the 4 demographic categories.

Malay Ghosh is Professor, Department of Statistics, University of Florida, Gainesville, FL 32611. Kannan Natarajan is Senior Research Biostatistician, Bristol-Myers Squibb Pharmaceutical Research Institute, Princeton, NJ 08543. T. W. F. Stroud is Professor, Department of Mathematics and Statistics, Queens University, Kingston, ON K7L 3N6, Canada. Bradley P. Carlin is Associate Professor, Division of Biostatistics, School of Public Health, University of Minnesota, Minneapolis, MN 55455. This research was supported in part by National Science Foundation grants SES-9201210 and SBR-9423996 (Ghosh), a grant from the Natural Science and Engineering Research Council of Canada (Stroud), and National Institute of Allergy and Infectious Diseases FIRST Award 1-R29-AI33466 and National Institute of Environmental Health Sciences grant 1-R01-ES07750 (Carlin). The authors thank James H. Albert of Bowling Green State University for many useful conversations, Robert Tsutakawa for his diligence in tracking down the county identifiers for the Missouri lung cancer dataset, and an associate editor and three referees for their thoughtful remarks, which led to substantial improvements in this article.

© 1998 American Statistical Association  
Journal of the American Statistical Association  
March 1998, Vol. 93, No. 441, Theory and Methods

Our HB cell probability estimates “borrow strength” from the other cells, resulting in smaller standard errors. Moreover, shrinkage toward the grand mean is done adaptively, in that the estimates reported for cells with larger sample sizes are shrunk less than those based on smaller sample sizes.

The second dataset relates to cancer mortality rates for the 115 counties in Missouri during 1972–1981. In each county, deaths due to lung cancer are broken down into four age groups (45–54, 55–64, 65–74, and 75+) and two sex groups (male and female). The number of deaths in some of these county subgroups during this period is very small (occasionally 0), so there is a clear need to borrow strength across cells. Tsutakawa (1988) and Tsutakawa, Shoop, and Marienfeld (1985) considered EB estimation of the rates for the given age groups, and Tsutakawa (1985) compared these EB rates with approximate Bayes rates, but these works dealt only with the male population and did not use prior distributions that could account for spatial similarity of the underlying rates in neighboring counties. We consider several possible models using such a spatial smoothing prior and including age, sex, and age–sex interaction as covariates. After selecting an appropriate model somewhat informally using a log-likelihood score statistic, we map the raw and fitted relative risks for a particular age–sex group as well as the fitted risks obtained in the earlier analysis by Tsutakawa, allowing the benefits of our spatial model to be assessed visually. We also investigate the adequacy of our model using a variety of model checks facilitated by our MCMC implementation.

## 2. HIERARCHICAL MODELS

Suppose that there are  $m$  strata or local areas. Let  $Y_{ik}$  denote the minimal sufficient statistic (discrete or continuous) for the  $k$ th unit within the  $i$ th stratum ( $k = 1, \dots, n_i; i = 1, \dots, m$ ). The  $Y_{ik}$  are assumed to be conditionally independent with pdf

$$f(y_{ik}|\theta_{ik}, \phi_{ik}) = \exp[\phi_{ik}^{-1}(y_{ik}\theta_{ik} - \psi(\theta_{ik})) + \rho(y_{ik}; \phi_{ik})] \quad (1)$$

( $k = 1, \dots, n_i; i = 1, \dots, m$ ). Such a model is referred to as a generalized linear model (McCullagh and Nelder 1989, p. 28). The density (1) is parameterized with respect to the canonical parameters  $\theta_{ik}$  and the scale parameters  $\phi_{ik}$  ( $> 0$ ). It is assumed that the scale parameters  $\phi_{ik}$  are known.

The natural parameters  $\theta_{ik}$  are first modeled as

$$h(\theta_{ik}) = \mathbf{x}_{ik}^T \boldsymbol{\beta} + u_i + \varepsilon_{ik} \quad (k = 1, \dots, n_i; i = 1, \dots, m), \quad (2)$$

where  $h$  is a strictly increasing function, the  $\mathbf{x}_{ik}$  ( $p \times 1$ ) are known design vectors,  $\boldsymbol{\beta}$  ( $p \times 1$ ) is the unknown regression coefficient,  $u_i$  are the random effects, and  $\varepsilon_{ik}$  are the errors. It is assumed that the  $u_i$  and the  $\varepsilon_{ik}$  are mutually independent with  $u_i \stackrel{\text{iid}}{\sim} N(0, \sigma_u^2)$  and  $\varepsilon_{ik} \stackrel{\text{iid}}{\sim} N(0, \sigma^2)$ .

It is possible to represent (1) and (2) in a hierarchical framework. Let  $R_u = \sigma_u^{-2}$  and  $R = \sigma^{-2}$ . Also, let  $\boldsymbol{\theta} = (\theta_{11}, \dots, \theta_{1n_1}, \dots, \theta_{m1}, \dots, \theta_{mn_m})^T$  and  $\mathbf{u} = (u_1, \dots, u_m)^T$ . Then the hierarchical model is given by the following:

- (I) Conditional on  $\boldsymbol{\theta}, \boldsymbol{\beta}, \mathbf{u}, R_u = r_u$ , and  $R = r$ ,  $Y_{ik}$  are independent with densities given in (1).
  - (II) Conditional on  $\boldsymbol{\beta}, \mathbf{u}, R_u = r_u$ , and  $R = r$ ,  $h(\theta_{ik}) \stackrel{\text{ind}}{\sim} N(\mathbf{x}_{ik}^T \boldsymbol{\beta} + u_i, r^{-1})$ .
  - (III) Conditional on  $\boldsymbol{\beta}, R_u = r_u$ , and  $R = r$ ,  $u_i \stackrel{\text{ind}}{\sim} N(0, r_u^{-1})$ .
- To complete the hierarchical model, we assign the following prior to  $\boldsymbol{\beta}, R_u = r_u$ , and  $R = r$ :
- (IV)  $\boldsymbol{\beta}, R_u = r_u$ , and  $R = r$  are mutually independent with  $\boldsymbol{\beta} \sim \text{uniform}(\mathbf{R}^p)$ , ( $p < m$ ),  $R_u \sim \text{gamma}(\frac{1}{2}a, \frac{1}{2}b)$ , and  $R \sim \text{gamma}(\frac{1}{2}c, \frac{1}{2}d)$ .

(A random variable  $Z \sim \text{gamma}(\alpha, \beta)$  if  $Z$  has pdf  $f(z) \propto \exp(-\alpha z) z^{\beta-1} I_{(0, \infty)}(z)$ .)

We are interested in finding the joint posterior distribution of the  $g(\theta_{ik})$ 's, where  $g$  is a strictly increasing function, given the data  $\mathbf{y} = (y_{11}, \dots, y_{1n_1}, \dots, y_{m1}, \dots, y_{mn_m})^T$ , and in particular in finding the posterior means, variances, and covariances of these parameters. In typical applications,  $g(\theta_{ik}) = \psi'(\theta_{ik}) = E(Y_{ik}|\theta_{ik})$ .

First, however, one must ensure that the joint posterior distribution of the  $\theta_{ik}$ 's given  $\mathbf{y}$  is proper. A theorem is proved to this effect. In what follows, the support of  $\theta_{ik}$  is the open interval  $(\theta_{ik}, \bar{\theta}_{ik})$ , where the lower endpoint of the interval can be  $-\infty$ , the upper endpoint can be  $+\infty$ , or both.

**Theorem 1.** Assume that  $a > 0, c > 0, \sum_i n_i - p + d > 0$ , and  $m + b > 0$ . Then, if

$$\int_{\theta_{ik}}^{\bar{\theta}_{ik}} \exp\{[\theta y_{ik} - \psi(\theta)]/\phi_{ik}\} h'(\theta) d\theta < \infty \quad (3)$$

for all  $y_{ik}$  and  $\phi_{ik}$  ( $> 0$ ), the joint posterior pdf of the  $\theta_{ik}$ 's given  $\mathbf{y}$  is proper.

The proof of this theorem is deferred to the Appendix.

Two special cases are of interest. In the first case,  $Y_{ik}|\theta_{ik} \sim \text{bin}(n_{ik}, \exp(\theta_{ik})/(1 + \exp(\theta_{ik})))$ . Suppose now that  $h$  is the identity function; that is, the link is canonical. Also, let  $g(\theta_{ik}) = \psi'(\theta_{ik})/n_{ik} = \exp(\theta_{ik})/[1 + \exp(\theta_{ik})]$ . Then, writing  $p_{ik} = \exp(\theta_{ik})/[1 + \exp(\theta_{ik})]$ , (3) reduces to  $\int_0^1 p_{ik}^{y_{ik}-1} (1 - p_{ik})^{n_{ik}-y_{ik}-1} dp_{ik} < \infty$ , which requires  $1 \leq y_{ik} \leq (n_{ik} - 1)$ ; that is, excludes cases of all failures or all successes. In the second case,  $Y_{ik}|\theta_{ik} \sim \text{Poisson}(\exp(\theta_{ik}))$ . Then, if  $h$  is the canonical link, and  $g(\theta_{ik}) = \psi'(\theta_{ik}) = \exp(\theta_{ik})$ , (3) reduces to  $\int_0^\infty \zeta_{ik}^{y_{ik}-1} \exp(-\zeta_{ik}) d\zeta_{ik} < \infty$ , which holds for  $y_{ik} = 1, 2, \dots$ . It may be noted, however, that although our general theorem needs this positivity restriction on the  $y_{ik}$  in the binomial and Poisson examples, recent work by Maiti (1997) showed that  $\sum_k y_{ik} > 0$  for each  $i$  is sufficient for posterior propriety.

Direct evaluation of the joint posterior distribution of the  $g(\theta_{ik})$ 's given  $\mathbf{y}$  involves high-dimensional numerical integration and is not computationally feasible. Instead, we use

the Gibbs sampler (Gelfand and Smith 1990; Geman and Geman 1984). Its implementation requires generating samples from certain conditional posterior distributions. Write  $\mathbf{h}(\boldsymbol{\theta}) = (h(\theta_{11}), \dots, h(\theta_{1n_1}), \dots, h(\theta_{m1}), \dots, h(\theta_{mn_m}))^T$ , and  $\mathbf{X} = (\mathbf{x}_{11}, \dots, \mathbf{x}_{1n_1}, \dots, \mathbf{x}_{m1}, \dots, \mathbf{x}_{mn_m})^T$ . Assume that  $\mathbf{X}^T \mathbf{X}$  is nonsingular. The necessary conditional distributions based on (I)–(IV) are

- (i)  $\beta | \boldsymbol{\theta}, \mathbf{u}, r_u, r, \mathbf{y} \sim N((\mathbf{X}^T \mathbf{X})^{-1}(\mathbf{X}^T \mathbf{h}(\boldsymbol{\theta}) - \sum_i u_i \sum_k \mathbf{x}_{ik}), r^{-1}(\mathbf{X}^T \mathbf{X})^{-1})$ ;
- (ii)  $u_i | \boldsymbol{\theta}, \beta, r_u, r, \mathbf{y} \stackrel{\text{ind}}{\sim} N((rn_i + r_u)^{-1} r \sum_k (h(\theta_{ik}) - \mathbf{x}_{ik}^T \beta), (rn_i + r_u)^{-1})$ ;
- (iii)  $R | \boldsymbol{\theta}, \beta, \mathbf{u}, r_u, \mathbf{y} \sim \text{gamma}(\frac{1}{2}(c + \sum_i \sum_k (h(\theta_{ik}) - \mathbf{x}_{ik}^T \beta - u_i)^2), \frac{1}{2}(d + \sum_i n_i))$ ;
- (iv)  $R_u | \boldsymbol{\theta}, \beta, \mathbf{u}, r, \mathbf{y} \sim \text{gamma}(\frac{1}{2}(a + \sum_i u_i^2), \frac{1}{2}(b + \sum_i n_i))$ ; and
- (v)  $\theta_{ik} | \beta, \mathbf{u}, r_u, r, \mathbf{y} \stackrel{\text{ind}}{\sim} \pi(\theta_{ik} | \beta, \mathbf{u}, r_u, r, \mathbf{y})$

$$\propto \exp \left[ (y_{ik} \theta_{ik} - \psi(\theta_{ik})) \phi_{ik}^{-1} - \frac{r}{2} (h(\theta_{ik}) - \mathbf{x}_{ik}^T \beta - u_i)^2 \right] h'(\theta_{ik}).$$

It is easy to generate samples from the normal and gamma distributions given in (i)–(iv). On the other hand, as evidenced in (v), the posterior distribution of  $\theta_{ik}$  given  $\beta, \mathbf{u}, r_u, r$ , and  $\mathbf{y}$  is known only up to a multiplicative constant, and accordingly one must use a general accept–reject algorithm to generate samples from this pdf. In the special case where  $h$  is the identity function, the task becomes much simpler due to the following lemma, which establishes log-concavity of  $\pi(\theta_{ik} | \beta, \mathbf{u}, r_u, r, \mathbf{y})$ . In such cases one can use the adaptive rejection sampling scheme of Gilks and Wild (1992).

**Lemma 1.** When  $h(z) = z$  for all  $z$ ,  $\log \pi(\theta_{ik} | \beta, \mathbf{u}, r, r_u, \mathbf{y})$  is a concave function of  $\theta_{ik}$ .

*Proof.* Straightforward.

Inference about  $\boldsymbol{\theta}$  will be based on (i)–(v). Indeed, based on (v), one can also find  $E(\theta_{ik} | \mathbf{y})$ ,  $V(\theta_{ik} | \mathbf{y})$ , and  $\text{cov}(\theta_{ik}, \theta_{i'k'} | \mathbf{y})(i, k) \neq (i', k')$  based on Monte Carlo integration techniques and formulas for iterated conditional expectations and variances.

The model considered in (I)–(IV) resembles closely the ones considered by Breslow and Clayton (1993), MacGibbon and Tomberlin (1989), and Zeger and Karim (1991). However, this model is not strictly contained in the one considered by Zeger and Karim (1991). Zeger and Karim considered  $h(\theta_{ik}) = \mathbf{x}_{ik}^T \beta + u_i$ , where  $h(\cdot)$  is a strictly increasing function, but this formulation does not include possible error in misspecifying this model. Indeed, according to our model, the uncertainty in specifying the model is broken up into two components: the effect of the local area and the error component. This allows the possibility of accounting for overdispersion by introducing an extra variance component.

Our method should also be contrasted to that of Albert (1988), which generalizes the approach of Leonard and

Novick (1986) and which was applied to binary survey data by Stroud (1994). Albert's method applied to the present setting first uses independent conjugate priors

$$\pi(\theta_{ik} | m_{ik}, \zeta) = \exp[\zeta(m_{ik} \theta_{ik} - \psi(\theta_{ik})) + g(m_{ik}; \zeta)] \quad (4)$$

for the  $\theta_{ik}$ . Next, it assumes that  $h(m_{ik}) = \mathbf{x}_{ik}^T \beta$  for some known monotone function  $h$ . Subsequently, it assigns distributions (possibly diffuse) to the hyperparameters  $\beta$  and  $\zeta$ . In contrast, our approach does not need the conjugacy of the prior and models monotone functions of  $\theta_{ik}$  instead of monotone functions of  $m_{ik} = E[\psi'(\theta_{ik})]$ . Moreover, Albert (1988) suggested approximation to the Bayes procedure by one of the following three methods: Laplace's method, quasi-likelihood approaches, or Brooks's (1984) method. These approximations generally are unnecessary now with the advent of the sophisticated MCMC integration techniques.

The log-concavity idea is used slightly differently by Dellaportas and Smith (1993), whose prime objective is inference about  $\beta$  in generalized linear models and model  $\theta_{ik}$  as functions of  $\beta$  without any error. In addition, their method, unlike ours, does not introduce any uncertainty in specifying the model.

We now examine how the previous results can be generalized for the analysis of multicategory data. Consider  $m$  strata labeled  $1, \dots, m$ . Within each stratum, several units are selected; suppose that the responses of individuals within each selected unit are independent and can be classified into  $J$  categories. For the  $k$ th selected unit within the  $i$ th stratum, let  $p_{ijk}$  denote the probability that an individual's response falls in the  $j$ th category ( $j = 1, \dots, J; k = 1, \dots, n_i$ ). Then within the  $k$ th selected unit within the  $i$ th stratum,  $Z_{ijk}$  ( $j = 1, \dots, J$ ) have a joint multinomial( $t_{ik}; p_{i1k}, \dots, p_{iJk}$ ) distribution, where  $t_{ik} = \sum_j Z_{ijk}$ . Using the well-known relationship between the multinomial and Poisson distributions,  $(Z_{i1k}, \dots, Z_{iJk})$  has the same distribution as the joint conditional distribution of  $(Y_{i1k}, \dots, Y_{iJk})$  given  $\sum_{j=1}^J Y_{ijk} = t_{ik}$ , where the  $Y_{ijk}$  ( $j = 1, \dots, J$ ) are independent Poisson( $\zeta_{ijk}$ ) and  $p_{ijk} = \zeta_{ijk} / \sum_{j=1}^J \zeta_{ijk}$  ( $j = 1, \dots, J$ ).

Let  $\theta_{ijk} = \log \zeta_{ijk}$ , and let  $\boldsymbol{\theta}$  denote the vector whose elements are the  $\theta_{ijk}$ 's. One can also model  $\theta_{ijk}$  as

$$h(\theta_{ijk}) = \mathbf{x}_{ijk}^T \beta + u_{ij} + \varepsilon_{ijk}. \quad (5)$$

Also, it is assumed that  $u_{ij}$  and the  $\varepsilon_{ijk}$  are mutually independent with  $u_{ij} \stackrel{\text{iid}}{\sim} N(0, \sigma_u^2)$  and  $\varepsilon_{ijk} \stackrel{\text{iid}}{\sim} N(0, \sigma^2)$ .

Then the hierarchical model, which is closely related to (I)–(IV), is given by the following:

- (A)  $Y_{ijk} | \boldsymbol{\theta}, \mathbf{u}, \beta, r_u, r$  are independent with

$$f(y_{ijk} | \boldsymbol{\theta}, \mathbf{u}, \beta, r_u, r) = \exp[\phi_{ijk}^{-1}(y_{ijk} \theta_{ijk} - \psi(\theta_{ijk})) + \rho(y_{ijk}; \phi_{ijk})].$$

- (B)  $h(\theta_{ijk}) | \mathbf{u}, \beta, r_u, r \stackrel{\text{ind}}{\sim} N(\mathbf{x}_{ijk}^T \beta + u_{ij}, r^{-1})$ .

- (C)  $u_{ij} | \beta, r_u, r \stackrel{\text{ind}}{\sim} N(0, r_u^{-1})$ .

- (D)  $\beta, R_u$ , and  $R$  are mutually independent with  $\beta \sim \text{uniform}(\mathbf{R}^p)$ ,  $R_u \sim \text{gamma}(\frac{1}{2}a, \frac{1}{2}b)$ , and  $R \sim \text{gamma}(\frac{1}{2}c, \frac{1}{2}d)$ .

We are interested in the posterior means, variances, and covariances of the  $p_{ijk} = \exp(\theta_{ijk}) / \sum_{j=1}^J \exp(\theta_{ijk})$  ( $k = 1, \dots, n_i; i = 1, \dots, m; j = 1, \dots, J$ ). The necessary posterior distributions for doing these calculations are given by

- (a)  $\beta | \theta, \mathbf{u}, r_u, r, \mathbf{y} \sim N((\sum_{i,j,k} \mathbf{x}_{ijk} \mathbf{x}_{ijk}^T)^{-1} (\sum_{i,j,k} \mathbf{x}_{ijk} (h(\theta_{ijk}) - u_{ij})), r^{-1} (\sum_{i,j,k} \mathbf{x}_{ijk} \mathbf{x}_{ijk}^T)^{-1})$ ;
- (b)  $u_{ij} | \theta, \beta, r_u, r, \mathbf{y} \stackrel{\text{ind}}{\sim} N((rn_i + r_u)^{-1} r \sum_k (h(\theta_{ijk}) - \mathbf{x}_{ijk}^T \beta), (rn_i + r_u)^{-1})$ ;
- (c)  $R | \theta, \beta, \mathbf{u}, r_u, \mathbf{y} \sim \text{gamma}(1/2(c + \sum_{i,j,k} (h(\theta_{ijk}) - \mathbf{x}_{ijk}^T \beta - u_{ij})^2), 1/2(d + \sum_i n_i))$ ;
- (d)  $R_u | \theta, \beta, \mathbf{u}, r, \mathbf{y} \sim \text{gamma}(1/2(a + \sum_i \sum_j u_{ij}^2), 1/2(b + mJ))$ ; and
- (e)  $\theta_{ijk} | \beta, \mathbf{u}, r_u, r, \mathbf{y} \stackrel{\text{ind}}{\sim} \pi(\theta_{ijk} | \beta, \mathbf{u}, r_u, r, \mathbf{y}) \propto \exp[(y_{ijk} \theta_{ijk} - \psi(\theta_{ijk})) \phi_{ijk}^{-1} - (r/2)(h(\theta_{ijk}) - \mathbf{x}_{ijk}^T \beta - u_{ij})^2] h'(\theta_{ijk})$ .

Once again posterior inference about  $g(\theta_{ijk})$ 's is performed using (e) and iterated formulas for posterior moments.

To conclude this section, we consider spatial HB GLMs and provide sufficient conditions for the propriety of the posterior. We begin with the likelihood given in (1) and model the  $\theta_{ik}$  as in (2), but this time the  $u_i$  represent variables that if observed would display spatial structure. More particularly, we model the  $u_i$  so that a pair of contiguous zones would have stronger (positive) correlation than any arbitrary zones that are noncontiguous.

For  $u_1, \dots, u_m$ , we consider the prior

$$p(u_1, \dots, u_m | r_u) \propto r_u^{m/2} \exp \left[ -\frac{r_u}{2} \sum_{i < l} w_{il} (u_i - u_l)^2 \right], \quad (6)$$

where the  $w_{il}$  are strictly positive if zones  $i$  and  $l$  are contiguous, and  $w_{il} = 0$  otherwise. This prior is a special case of general pairwise difference priors, considered by, for example, Besag, Green, Higdon, and Mengersen (1995). The marginal priors for  $\beta$ ,  $R_u$ , and  $R$  remain as before.

For brevity, write  $n_T = \sum_{i=1}^m n_i$  and  $\bar{\mathbf{x}} = n_T^{-1} \sum_{i=1}^m \sum_{k=1}^{n_i} \mathbf{x}_{ik}$ . It is assumed that the matrix  $\mathbf{X}_0^T = (\mathbf{x}_{11} - \bar{\mathbf{x}}, \dots, \mathbf{x}_{1n_1} - \bar{\mathbf{x}}, \dots, \mathbf{x}_{m1} - \bar{\mathbf{x}}, \dots, \mathbf{x}_{mn_m} - \bar{\mathbf{x}})$  has rank  $p$ . We then obtain the following theorem.

**Theorem 2.** Assume the conditions of Theorem 1, but where now  $n_T - p + d > 1$ . Then the joint posterior of the  $\theta_{ik}$  under the spatial prior (6) is proper.

The proof of this theorem is also deferred to the Appendix. For implementing this Bayes procedure via Gibbs sampling, one finds conditional distributions similar to (i)–(v) earlier, with minor modifications to (ii) and (iv).

**Remark.** It should be noted that if instead of (2), one models the  $\theta_{ik}$  as

$$h(\theta_{ik}) = \beta_0 + \mathbf{x}_{ik}^T \beta + u_i + \varepsilon_{ik} \\ (k = 1, \dots, n_i; i = 1, \dots, m),$$

then the posterior of the  $h(\theta_{ik})$  fails to be proper. The introduction of the intercept term  $\beta_0$  creates a nonidentifiability in the posterior, which in turn implies that the joint posterior of the  $g(\theta_{ik})$  is also improper.

### 3. DATA ANALYSIS

#### 3.1 Exposure to Health Hazards Dataset

The analysis of the multicategory dataset mentioned in Section 1, where persons in 15 regions of Canada were asked the question "Have you experienced any negative impact of exposure to health hazards in the workplace," is reported in Table 1 and Figure 1. Here for the  $k$ th selected age–sex category within the  $i$ th region,  $p_{ijk}$  denotes the probability that an individual's response falls in the  $j$ th category (where the categories are 1 = yes, 2 = no, 3 = not exposed, and 4 = not applicable or not stated). Within the  $k$ th selected age–sex category and the  $i$ th region, the  $Z_{ijk}$  have a joint multinomial( $t_{ik}; p_{i1k}, \dots, p_{iJk}$ ) distribution, where  $t_{ik} = \sum_j Z_{ijk}$ . Fitting model (5) with the Poisson likelihood as described in Section 2, and relabeling  $k$  as  $(a, s)$  for clarity, the regression equation is

$$\mathbf{x}_{ijk}^T \beta = \mu + \tau_a^A + \tau_s^S + \tau_j^J + \tau_{as}^{AS} + \tau_{aj}^{AJ} + \tau_{sj}^{SJ},$$

where  $\mu$  is the general effect,  $\tau_a^A$  is the main effect due to the  $a$ th age group,  $\tau_s^S$  is the main effect due to the  $s$ th sex,  $\tau_j^J$  is the main effect due to the  $j$ th category response,  $\tau_{as}^{AS}$  is the interaction effect of the  $a$ th age and  $s$ th sex,  $\tau_{aj}^{AJ}$  is the interaction effect of the  $a$ th age and  $j$ th category response, and  $\tau_{sj}^{SJ}$  is the interaction effect  $s$ th sex and  $j$ th category response. To avoid redundancy, we assume the corner point restrictions

$$\begin{aligned} \tau_1^A &= \tau_1^S = \tau_1^J = \tau_{a1}^{AS} = \tau_{1s}^{AS} \\ &= \tau_{a1}^{AJ} = \tau_{1j}^{AJ} = \tau_{s1}^{SJ} = \tau_{1j}^{SJ} = 0 \end{aligned}$$

for all  $a, s$ , and  $j$ .

Using the extremely vague (but proper) priors for  $R_u$  and  $R$  determined by setting  $a = b = c = d = .002$ , we generated 10 parallel Gibbs sampling chains of 2,000 iterations each. Using the 1,000 samples from the latter half of these chains (iterations 1,001–2,000), Table 1 contains the HB estimates, the sample proportions, and the associated standard errors for all four categories in each of the cells cross-classified by  $2 \times 2 = 4$  demographic categories for three regions: the smallest, the median, and the largest. Figure 1 shows the sample proportions ("Prop"), traditional logistic regression estimates ("Regr"), and hierarchical Bayes estimates "HB") for all 15 regions for females age 40 or younger. For regions with larger overall sample sizes, shrinkage of the estimates toward the logistic regression estimates within each age–sex category is much smaller than that observed in the smaller regions. For example, Figure 1b shows the HB estimates to be very similar to the logistic regression estimates in the sparsely populated Region 2, whereas Figure 1h shows HB estimates very much like the original sample proportions in populous Region 8. Also, within the  $k$ th age–sex category in

Table 1. Impact of Exposure to Health Hazards in the Workplace

| Category    | Response    | Sample         |      | H. Bayes    |      |
|-------------|-------------|----------------|------|-------------|------|
|             |             | Proportions    | SD   | Proportions | SD   |
| Region = 2  |             | Total n = 294  |      |             |      |
| M, Age < 40 | Yes         | .400           | .100 | .373        | .042 |
|             | No          | .383           | .101 | .345        | .041 |
|             | Not exposed | .150           | .119 | .199        | .031 |
|             | NA/NS       | .067           | .125 | .083        | .015 |
| F, Age < 40 | Yes         | .257           | .100 | .266        | .035 |
|             | No          | .284           | .098 | .279        | .035 |
|             | Not exposed | .311           | .097 | .274        | .036 |
|             | NA/NS       | .148           | .107 | .181        | .026 |
| M, Age ≥ 40 | Yes         | .111           | .111 | .184        | .028 |
|             | No          | .153           | .109 | .176        | .027 |
|             | Not exposed | .167           | .108 | .156        | .026 |
|             | NA/NS       | .569           | .077 | .484        | .040 |
| F, Age ≥ 40 | Yes         | .159           | .098 | .110        | .019 |
|             | No          | .091           | .102 | .103        | .018 |
|             | Not exposed | .125           | .010 | .134        | .022 |
|             | NA/NS       | .625           | .065 | .654        | .034 |
| Region = 3  |             | Total n = 740  |      |             |      |
| M, Age < 40 | Yes         | .294           | .070 | .311        | .029 |
|             | No          | .426           | .063 | .395        | .032 |
|             | Not exposed | .203           | .075 | .186        | .023 |
|             | NA/NS       | .077           | .080 | .108        | .015 |
| F, Age < 40 | Yes         | .246           | .064 | .235        | .024 |
|             | No          | .273           | .063 | .287        | .026 |
|             | Not exposed | .180           | .067 | .204        | .023 |
|             | NA/NS       | .301           | .062 | .274        | .026 |
| M, Age ≥ 40 | Yes         | .156           | .069 | .154        | .019 |
|             | No          | .150           | .069 | .165        | .020 |
|             | Not exposed | .100           | .071 | .112        | .016 |
|             | NA/NS       | .594           | .048 | .569        | .028 |
| F, Age ≥ 40 | Yes         | .064           | .063 | .071        | .010 |
|             | No          | .086           | .063 | .091        | .012 |
|             | Not exposed | .111           | .062 | .099        | .013 |
|             | NA/NS       | .739           | .033 | .739        | .021 |
| Region = 8  |             | Total n = 1707 |      |             |      |
| M, Age < 40 | Yes         | .274           | .047 | .279        | .021 |
|             | No          | .360           | .044 | .362        | .023 |
|             | Not exposed | .253           | .048 | .253        | .020 |
|             | NA/NS       | .113           | .052 | .106        | .012 |
| F, Age < 40 | Yes         | .199           | .042 | .196        | .016 |
|             | No          | .267           | .040 | .275        | .019 |
|             | Not exposed | .289           | .040 | .295        | .019 |
|             | NA/NS       | .245           | .041 | .234        | .017 |
| M, Age ≥ 40 | Yes         | .113           | .047 | .130        | .013 |
|             | No          | .166           | .046 | .174        | .016 |
|             | Not exposed | .217           | .044 | .195        | .017 |
|             | NA/NS       | .504           | .035 | .501        | .022 |
| F, Age ≥ 40 | Yes         | .087           | .042 | .076        | .009 |
|             | No          | .123           | .041 | .110        | .011 |
|             | Not exposed | .119           | .041 | .131        | .012 |
|             | NA/NS       | .671           | .025 | .683        | .017 |

the  $i$ th region, the shrinkage is again smaller for categories with larger numbers of responses. For example, of females over age 40 in Region 3, 15/234 (6.41%) answered “yes,” compared to 173/234 (73.93%) in the “not applicable/not stated” category. As seen in Table 1, the shrinkage is much larger for the former case, again revealing the adaptive nature of the HB estimates. Finally, note that the standard errors associated with the HB estimates are much smaller than those for the sample proportions.

### 3.2 Missouri Lung Cancer Dataset

Our second example relates to lung cancer mortality rates in the 115 counties in Missouri during the period 1972–1981. Following the original analysis of this data by Tsutakawa (1985, 1988), we separate the city of St. Louis from the remainder of St. Louis County, which surrounds it. Mortality was classified for each county by sex into four age groups: 45–54, 55–64, 65–74, and 75 and older. The population size for each cell was taken to be the midperiod pop-

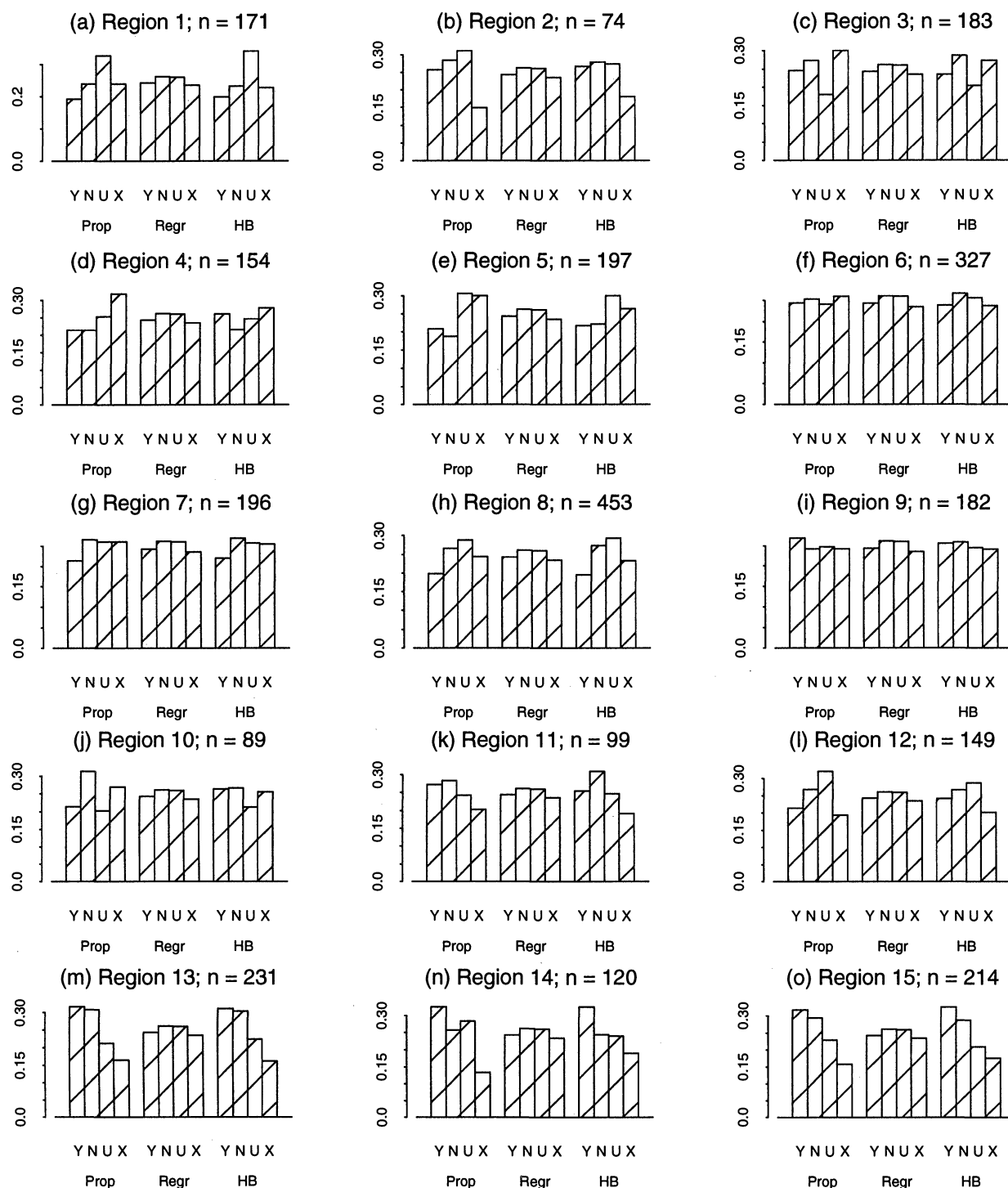


Figure 1. Bar Graph of Estimated Proportions by Category by Region for Females Age < 40. Prop = sample proportion, Regr = logistic regression estimate, HB = hierarchical Bayes estimate. Categories: Y = Yes, N = No, U = Not exposed, X = Not applicable or not stated.

ulation, obtained from the 1970 and 1980 U.S. censuses by linear interpolation.

Again relabeling  $k$  as  $(a, s)$  for clarity, let  $Y_{ias}$  be the lung cancer death count and  $n_{ias}$  the midperiod population in the  $i$ th county for the  $a$ th age group and  $s$ th sex,  $i = 1, \dots, 115$ ,  $a = 1, \dots, 4$ ,  $s = 1, 2$ . At the first stage of the model, we assume that  $Y_{ias} | \zeta_{ias} \stackrel{\text{ind}}{\sim} \text{Poisson}(\zeta_{ias})$ . We then model the mean structure by assuming that  $\zeta_{ias} = E_{ias} \exp(\mu_{ias})$ , where  $E_{ias}$  is the number of deaths that would be expected using some current reference standard

and  $\mu_{ias}$  is the corresponding log-relative risk in cell  $ias$ . Some spatial analyses (see, e.g., Bernardinelli and Montomoli 1992) have used an externally available reference table to compute the  $E_{ias}$ ; here we adopt the simpler alternative of *internal* standardization, defining  $E_{ias} = n_{ias} \cdot r$ , where  $r \equiv \sum_{ias} Y_{ias} / \sum_{ias} n_{ias}$ , the statewide lung cancer rate over all sex and age groups in our dataset.

The log-relative risks are then modeled linearly as

$$\mu_{ias} = \mathbf{x}_{as}^T \boldsymbol{\beta} + u_i + \varepsilon_{ias}, \quad (7)$$

Table 2. Informal Model Comparison, Missouri Lung Cancer Data

| Model for $\mathbf{x}_{as}^T \beta$                                                                                                                                   | Number of fixed effects | Log-likelihood score, $\bar{l}$ | Difference |
|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------|-------------------------|---------------------------------|------------|
| $v_s \alpha + z_a \gamma$                                                                                                                                             | 2                       | 580.0                           |            |
| $v_s \alpha + z_a \gamma + v_s z_a \xi$                                                                                                                               | 3                       | 597.8                           | 17.8       |
| $v_s \alpha + z_a^{(L)} \gamma^{(L)} + v_s z_a^{(L)} \xi^{(L)} + z_a^{(U)} \gamma^{(U)} + v_s z_a^{(U)} \xi^{(U)}$                                                    | 5                       | 614.9                           | 17.1       |
| $v_s \alpha + z_a^{(L)} \gamma^{(L)} + v_s z_a^{(L)} \xi^{(L)} + z_a^{(M)} \gamma^{(M)} + v_s z_a^{(M)} \xi^{(M)} + z_a^{(U)} \gamma^{(U)} + v_s z_a^{(U)} \xi^{(U)}$ | 7                       | 618.7                           | 3.8        |

where  $\beta$  is a vector parameter that captures the effect of sex, age, and sex–age interaction. The  $\varepsilon_{ias}$  are assumed iid  $N(0, \sigma^2)$ , but the  $u_i$  account for potential spatial clustering of the rates via a conditionally autoregressive (CAR) prior structure (see, e.g., Besag, York, and Mollié 1991; Clayton and Kaldor 1987). That is, we assume that

$$u_i | u_{l \neq i} \sim N(\bar{u}_i, 1/(\tau m_i)),$$

where  $\bar{u}_i$  is the average of the  $u_{l \neq i}$  that are defined to be “neighbors” of  $u_i$ , and  $m_i$  is the number of these neighbors. Here we adopt the most common implementation of the CAR structure, defining two counties to be neighbors if and only if they are physically adjacent to each other. It is easy to show that this prior is of the form given in (6), where  $w_{il} = 1$  if counties  $i$  and  $l$  are adjacent, and 0 otherwise. Note that this CAR prior is defined only up to additive constant, again explaining the lack of an intercept term in (7).

It thus remains to determine the appropriate structure for  $\beta$ . Tsutakawa (1988) noted a strong similarity between the male death rates in the two oldest age groups, perhaps due to the competing risks of other diseases. Preliminary analysis of the female rates suggests a similar situation, and as such we begin by defining the sex and age scores

$$v_s = \begin{cases} 0 & \text{if } s = 1 \text{ (male)} \\ 1 & \text{if } s = 2 \text{ (female)} \end{cases}$$

and

$$z_a = \begin{cases} -1 & \text{if } a = 1 \text{ (age 45–54)} \\ 0 & \text{if } a = 2 \text{ (age 55–64)} \\ 1 & \text{if } a = 3 \text{ (age 65–74)} \\ 1 & \text{if } a = 4 \text{ (age 75+)} \end{cases}$$

and use them in a regression-type model,

$$\mathbf{x}_{as}^T \beta = v_s \alpha + z_a \gamma + v_s z_a \xi,$$

thus effectively combining the two oldest age groups. We complete our model specification with flat priors on the components of the fixed effect vector  $\beta$ , a vague  $\text{gamma}(.01, .01)$  hyperprior on  $\tau$ , and a moderately informative  $\text{gamma}(1, 1)$  hyperprior on  $R = 1/\sigma^2$ . (This latter hyperprior ensures a well-identified joint posterior distribution and, as we shall see, is still quite vague relative to the posterior for the  $\varepsilon_{ias}$ .) We then fit this model via Gibbs sampling using the BUGS language (Spiegelhalter, Thomas,

Best, and Gilks 1995), aided by the CODA S+ function (Best, Cowles, and Vines 1995) for assessing convergence and computing posterior summaries. BUGS uses S-like syntax for specifying fairly complex hierarchical models. The program converts this syntax into a directed acyclic graph, the nodes of which correspond to the complete conditional distributions necessary for the Gibbs algorithm. Our results

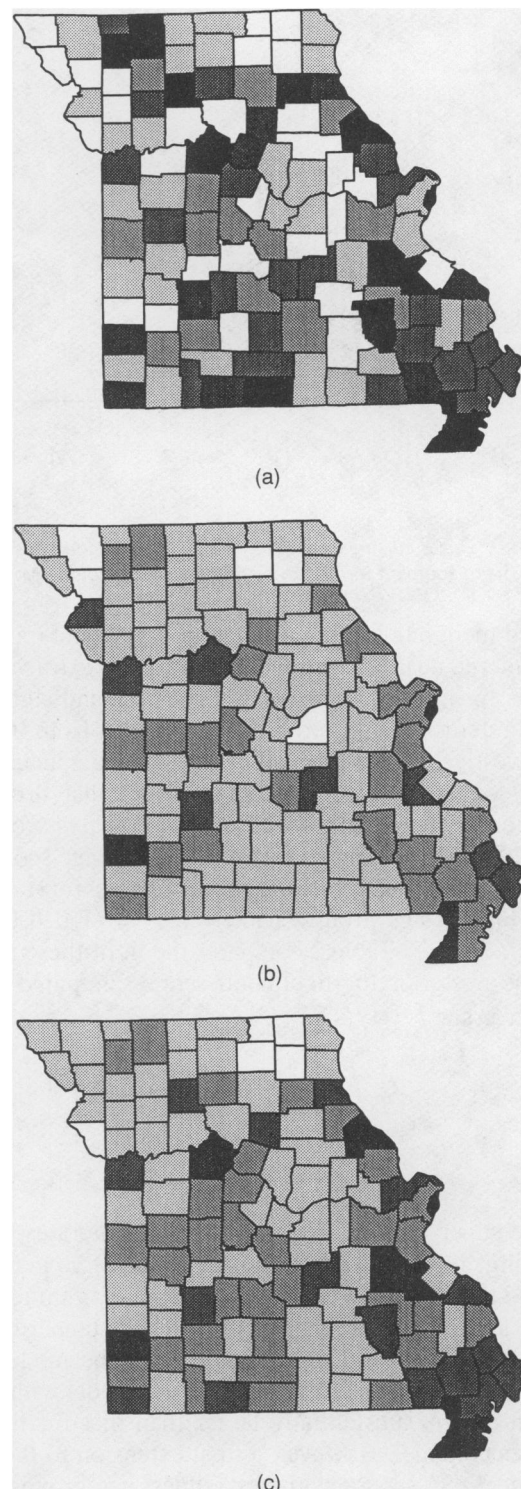


Figure 2. Male 55–64 Lung Cancer Relative Risks, Missouri Counties, 1972–1981. (a) Raw SMRs; (b) Tsutakawa EB smoothed RRs; (c) Spatially smoothed RRs.

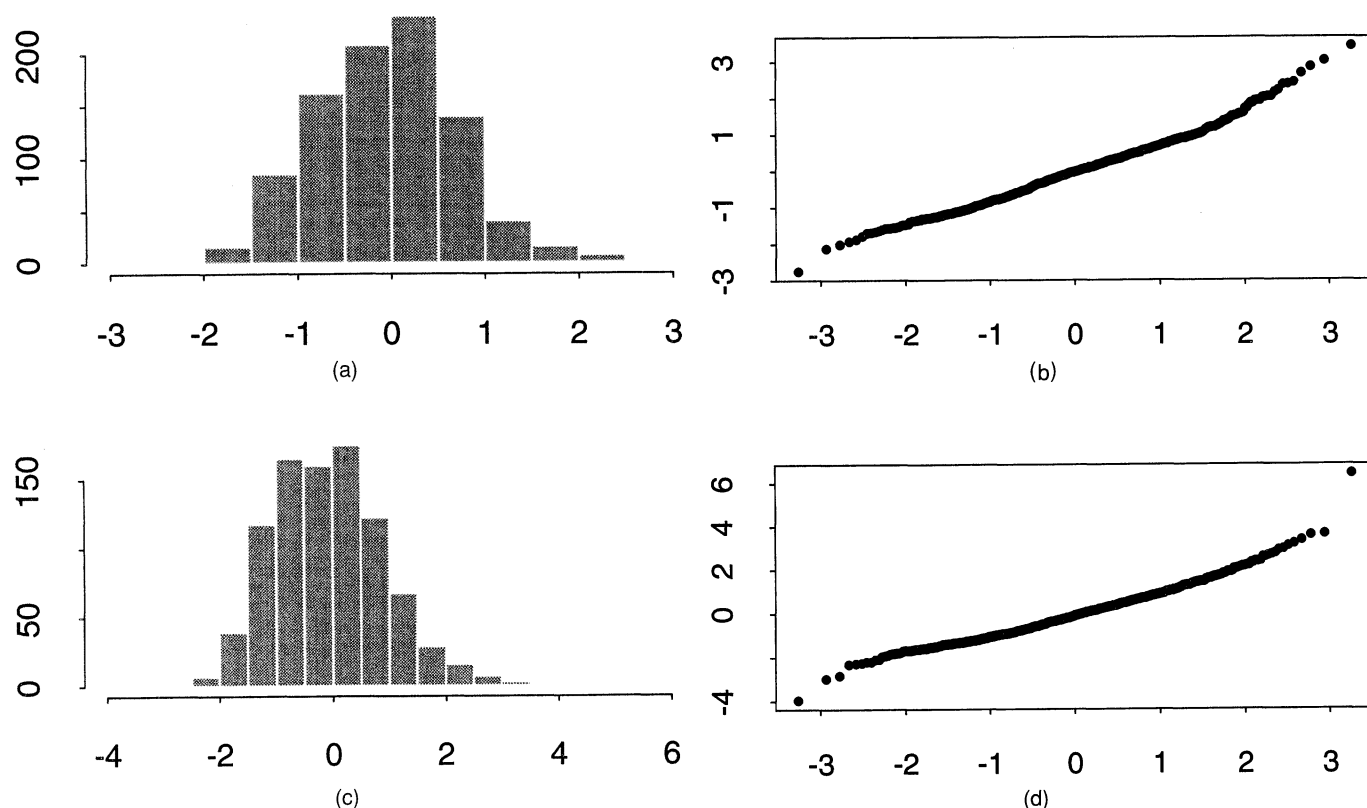


Figure 3. Residual Analysis, Five Fixed-Effects Model, Missouri Lung Cancer Data. (a) Histogram, with overdispersion terms; (b) normal q-q plot, with overdispersion terms; (c) histogram, no overdispersion terms; (d) normal q-q plot, no overdispersion terms.

indicated moderate spatial correlation in the data (posterior for  $\tau$  centered near 30), a modest need for the extra variability terms (posterior for  $\sigma$  centered near .2), and significant sex-age interaction (posterior for  $\xi$  removed from 0).

To investigate the scope of models that our data could support, we considered a simpler model that drops the interaction term  $\xi$  and a more complex analysis of variance (ANOVA)-type model that replaced the age score vector  $\mathbf{z} = (-1, 0, 1, 1)'$  with separate vectors for the lowest and highest age groups, namely  $\mathbf{z}^{(L)} = (1, 0, 0, 0)'$  and  $\mathbf{z}^{(U)} = (0, 0, 1, 1)'$ . Table 2 compares the fit of these models using the posterior log-likelihood score, computed as the sample average  $\bar{l} = 1/G \sum_{g=1}^G l^{(g)}$ , where

$$l^{(g)} = \sum_{ias} \mu_{ias}^{(g)} y_{ias} - \sum_{ias} E_{ias} \exp(\mu_{ias}^{(g)}) + C,$$

$$g = 1, \dots, G.$$

Here the superscript  $(g)$  indexes the Gibbs iterates, and  $C$  is a scaling constant. After a burn-in period of 50 iterations, we found that retaining  $G = 500$  iterations was sufficient to produce log-likelihood scores with batched standard errors near .5. Note that the average score  $\bar{l}$  for the model with five fixed effects is larger than that for the model with three, which in turn is substantially larger than that for the two fixed-effects model. However, a final extension to the saturated model that separates the two oldest age groups—that is, using  $\mathbf{z}^{(M)} = (0, 0, 1, 0)'$  and  $\mathbf{z}^{(U)} = (0, 0, 0, 1)'$ —offers no numerically significant improvement in fit. Although the usual chi-squared asymptotics for differences in  $-2\bar{l}$  are not

appropriate in our Bayesian random-effects model setting, it seems clear from Table 2 that the model with five fixed effects offers the best fit while preserving parsimony.

Our chosen model produces posterior means and 95% equal-tail credible sets as follows: for  $\alpha$ ,  $-1.46$  and  $(-1.545, -1.36)$ ; for  $\gamma^{(L)}$ ,  $-1.064$  and  $(-1.15, -.976)$ ; for  $\gamma^{(U)}$ ,  $.558$  and  $(.503, .630)$ ; for  $\xi^{(L)}$ ,  $.369$  and  $(.227, .503)$ ; and for  $\xi^{(U)}$ ,  $-.318$  and  $(-.428, -.207)$ . Thus log-relative risk is nearly 1.5 units lower for females than for males on average, with the risk increasing monotonically with age. (Recall that the two oldest age groups have been combined.) However, the signs on the interaction terms  $\xi^{(L)}$  and  $\xi^{(U)}$  show that this increase is not as dramatic for females as for males.

Figure 2 maps the raw standardized mortality ratios for men age 55–64,  $SMR_{i21} = Y_{i21}/E_{i21}$ , the fitted relative risks obtained by Tsutakawa (1988) using EB methods without a spatial smoothing prior, and the fitted relative risks from our fully Bayesian spatial smoothing analysis,  $RR_{i21} = 1/G \sum_{g=1}^G \exp(u_i^{(g)} + \varepsilon_{i21}^{(g)})$ , the average of the  $G = 500$  corresponding postconvergence relative risk estimates. Although the comparison between our results and Tsutakawa's is not completely fair, because the latter were obtained using data for males only, clearly both of these methods eliminate much of the noise in the original map while preserving the high rate in populous St. Louis city. However, our spatial model clarifies the general increase in rates from north to south (especially along the eastern border with Illinois) and also identifies possible clusters of counties with similar risk, while maintaining a reasonable amount of fidelity to the original data.

Finally, we check our model by analyzing the posterior means of the collection of standardized residuals,  $r_{ias} = E[(Y_{ias} - \zeta_{ias})/\sqrt{\zeta_{ias}}|\mathbf{y}]$ , which are readily computable in BUGS (Spiegelhalter et al. 1995, pp. 40–46). Figure 3a shows a histogram of these mean residuals, and Figure 3b gives their normal Q-Q plot. Both plots reveal a high degree of normality.

Finally, the rather small fitted standard deviation (.17) for the extra viability terms  $\varepsilon_{ias}$  made us wonder whether these terms were even needed in the model. To check this, we reran our model without these terms, obtaining the residual histogram and normal plot shown in Figures 3c and 3d. Although the degree of normality is still acceptable, the presence of a few large outliers is disturbing. The one enormous outlier on the high side corresponds to men in the youngest age group who live in the city of St. Louis; apparently their very high lung cancer death rate is poorly fit by the model. Interestingly, two of the three outlying values on the low side are the youngest and second-youngest groups of men living in St. Louis County, who are apparently much healthier than the model predicts. Thus we conclude that the overdispersion terms  $\varepsilon_{ias}$  are critical in obtaining acceptable fits in all-urban St. Louis city and its only geographic neighbor, suburban St. Louis County, allowing differing rates in these two disparate regions despite their juxtaposition on the map.

#### 4. CONCLUSIONS

In this article we have provided a general approach for small-area estimation based on hierarchical Bayes generalized linear models, with and without spatial correlation structure. Sufficient conditions have been given to ensure the propriety of posteriors under noninformative priors. The general methodology is applicable to a wide variety of situations calling for simultaneous estimation of small-area parameters. Future work looks to continued automation in the fitting of these models via MCMC methods, especially in the areas of model choice and model averaging. Promising tools in this regard include expected predicted deviance scores, recently introduced by Gelfand and Ghosh (1997) and illustrated for spatio-temporal models by Waller, Carlin, Xia, and Gelfand (1997).

#### APPENDIX: PROOFS

##### Proof of Theorem 1

The joint posterior pdf of  $\theta, \beta, \mathbf{u}, R_u$ , and  $R$  given  $\mathbf{y}$  is

$$\begin{aligned} \pi(\theta, \beta, \mathbf{u}, r_u, r|\mathbf{y}) &\propto \prod_i \prod_k \exp[\phi_{ik}^{-1}(y_{ik}\theta_{ik} - \psi(\theta_{ik}))] r^{1/2} \sum_1^m n_i \\ &\times \prod_i \prod_k \exp\left[-\frac{r}{2} (h(\theta_{ik}) - \mathbf{x}_{ik}^T \beta - u_i)^2\right] \\ &\times \left(\prod_i \prod_k h'(\theta_{ik})\right) r_u^{m/2} \exp\left(-\frac{r_u}{2} \sum_1^m u_i^2\right) \\ &\times \exp\left(-\frac{ar_u}{2}\right) r_u^{1/2b-1} \exp\left(-\frac{cr}{2}\right) r^{1/2d-1}. \end{aligned}$$

Integrating with respect to  $\beta, r_u$ , and  $r$  in succession, we obtain

$$\begin{aligned} \pi(\theta, \mathbf{u}|\mathbf{y}) &\leq C \prod_i \prod_k \exp[\phi_{ik}^{-1}(y_{ik}\theta_{ik} - \psi(\theta_{ik}))] \\ &\times \left(a + \sum_1^m u_i^2\right)^{-1/2(m+b)} \prod_i \prod_k h'(\theta_{ik}), \end{aligned}$$

where  $C (> 0)$  is a generic constant that does not depend on  $\theta$  or  $\mathbf{u}$ . Now integrating with respect to  $\mathbf{u}$  and using the structure of a multivariate  $t$ , it follows that

$$\pi(\theta|\mathbf{y}) \leq C \prod_i \prod_k \exp[\phi_{ik}^{-1}(y_{ik}\theta_{ik} - \psi(\theta_{ik}))] h'(\theta_{ik}).$$

The result now follows from (3).

##### Proof of Theorem 2

For notational simplicity, without loss of generality  $h$  is taken as the identity function throughout. The joint posterior of  $\theta, \beta, \mathbf{u}, R_u$ , and  $R$  given  $\mathbf{y}$  is

$$\begin{aligned} \pi(\theta, \beta, \mathbf{u}, r_u, r|\mathbf{y}) &\propto \prod_i \prod_k \exp[\phi_{ik}^{-1}(y_{ik}\theta_{ik} - \psi(\theta_{ik}))] r^{n_T/2} \\ &\times \prod_i \prod_k \exp\left[-\frac{r}{2} (\theta_{ik} - \mathbf{x}_{ik}^T \beta - u_i)^2\right] \\ &\times r_u^{m/2} \exp\left[-\frac{r_u}{2} \sum_{1 \leq i < l \leq m} w_{il} (u_i - u_l)^2\right] \\ &\times \exp\left(-\frac{ar_u}{2}\right) r_u^{1/2b-1} \exp\left(-\frac{cr}{2}\right) r^{1/2d-1}. \end{aligned}$$

With the one-to-one transformation  $(z_1, \dots, z_{m-1}, u_m)$ , where  $z_i = u_i - u_m, i = 1, \dots, m$ , the posterior transforms to

$$\begin{aligned} \pi(\theta, \beta, u_m, \mathbf{z}, r_u, r|\mathbf{y}) &\propto \exp\left[\sum_i \sum_k \phi_{ik}^{-1}(y_{ik}\theta_{ik} - \psi(\theta_{ik}))\right] r^{n_T/2} \\ &\times \exp\left[-\frac{r}{2} \sum_i \sum_k (\theta_{ik} - \mathbf{x}_{ik}^T \beta - z_i - u_m)^2\right] \\ &\times \exp\left[-\frac{r_u}{2} \left\{a + \sum_{1 \leq i < l \leq m} w_{il} (z_i - z_l)^2\right\}\right] r_u^{(m+b-1)/2} \\ &\times \exp\left(-\frac{cr}{2}\right) r^{1/2d-1}, \end{aligned}$$

where  $z_m = 0$  and  $\mathbf{z} = (z_1, \dots, z_{m-1})$ . Next, write  $\bar{\theta} = n_T^{-1} \sum_i \sum_k \theta_{ik}$  and  $\bar{z} = m^{-1} \sum_i z_i$ . Integrating with respect to  $u_m, \beta, r_u$ , and  $r$  in succession, we have

$$\begin{aligned} \pi(\theta, \mathbf{z}|\mathbf{y}) &\leq C \exp\left[\sum_i \sum_k \phi_{ik}^{-1}(y_{ik}\theta_{ik} - \psi(\theta_{ik}))\right] \\ &\times \left[a + \sum_{1 \leq i < l \leq m} w_{il} (z_i - z_l)^2\right]^{-(m+b)/2}, \end{aligned}$$

where  $C (> 0)$  is a generic constant that does not depend on  $\theta$  or  $\mathbf{z}$ . Recall that  $z_m = 0$  and  $\sum_{1 \leq i < l \leq m} w_{il} (z_i - z_l)^2$  involves only  $m$

— 1 variables  $z_1, \dots, z_{m-1}$ . Thus, integrating with respect to  $\mathbf{z}$ , and using the structure of a multivariate  $t$  distribution yields

$$\pi(\theta|\mathbf{y}) \leq C \exp \left[ \sum_i \sum_k \phi_{ik}^{-1} (y_{ik}\theta_{ik} - \psi(\theta_{ik})) \right].$$

The result again follows from (3).

[Received December 1995. Revised June 1997.]

## REFERENCES

- Albert, J. H. (1988), "Computational Methods Using a Bayesian Hierarchical Generalized Linear Model," *Journal of the American Statistical Association*, 83, 1037–1044.
- Bernardinelli, L., and Montomoli, C. (1992), "Empirical Bayes Versus Fully Bayesian Analysis of Geographical Variation in Disease Risk," *Statistics in Medicine*, 11, 983–1007.
- Besag, J., Green, P., Higdon, D., and Mengersen, K. (1995), "Bayesian Computation and Stochastic Systems" (with discussion), *Statistical Science*, 10, 3–66.
- Besag, J., York, J. C., and Mollié, A. (1991), "Bayesian Image Restoration, With Two Applications in Spatial Statistics" (with discussion), *Annals of the Institute of Statistical Mathematics*, 43, 1–59.
- Best, N. G., Cowles, M. K., and Vines, K. (1995), "CODA: Convergence Diagnosis and Output Analysis Software for Gibbs Sampling Output, Version 0.30," technical report, Cambridge University, Institute of Public Health, Medical Research Council Biostatistics Unit.
- Breslow, N. E., and Clayton, D. G. (1993), "Approximate Inference in Generalized Linear Mixed Models," *Journal of the American Statistical Association*, 88, 9–25.
- Brooks, R. J. (1984), "Approximate Likelihood Ratio Tests in the Analysis of Beta-Binomial Data," *Applied Statistics*, 33, 285–289.
- Clayton, D. G., and Kaldor, J. (1987), "Empirical Bayes Estimates of Age-Standardized Relative Risks for Use in Disease Mapping," *Biometrics*, 43, 671–681.
- Datta, G. S., and Ghosh, M. (1991), "Bayesian Prediction in Linear Models: Applications to Small Area Estimation," *The Annals of Statistics*, 19, 1748–1770.
- Dellaportas, P., and Smith, A. F. M. (1993), "Bayesian Inference for Generalized Linear and Proportional Hazards Models via Gibbs Sampling," *Applied Statistics*, 42, 443–459.
- Dempster, A. P., and Tomberlin, T. J. (1980), "The Analysis of Census Undercount From a Post-Enumeration Survey," in *Proceedings of the Conference on Census Undercount*, pp. 88–94.
- Farrell, P., MacGibbon, B., and Tomberlin, T. J. (in press), "Empirical Bayes Estimators of Small Area Proportions in Multistage Designs," submitted to *Statistica Sinica*.
- Fay, R. E., and Herriot, R. (1979), "Estimates of Income for Small Places: An Application of James–Stein Procedures to Census Data," *Journal of the American Statistical Association*, 74, 269–277.
- Gelfand, A. E., and Ghosh, S. K. (in press), "Model Choice: A Minimum Posterior Predictive Loss Approach," submitted to *Biometrika*.
- Gelfand, A. E., and Smith, A. F. M. (1990), "Sampling-Based Approaches to Calculating Marginal Densities," *Journal of the American Statistical Association*, 85, 398–409.
- Gelman, A., and Rubin, D. B. (1992), "Inference From Iterative Simulation Using Multiple Sequences," *Statistical Science*, 7, 457–511.
- Geman, S., and Geman, D. (1984), "Stochastic Relaxation, Gibbs Distributions and the Bayesian Restoration of Images," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 6, 721–741.
- Ghosh, M., and Lahiri, P. (1987), "Robust Empirical Bayes Estimation of Means From Stratified Samples," *Journal of the American Statistical Association*, 82, 1153–1162.
- (1992), "A Hierarchical Bayes Approach to Small Area Estimation With Auxiliary Information," in *Bayesian Analysis in Statistics and Econometrics*, Lecture Notes in Statistics 75, eds. P. K. Goel and N. S. Iyengar, New York: Springer-Verlag, pp. 107–125.
- Ghosh, M., and Meeden, G. (1986), "Empirical Bayes Estimation in Finite Population Sampling," *Journal of the American Statistical Association*, 74, 269–277.
- Ghosh, M., and Rao, J. N. K. (1994), "Small Area Estimation: An Appraisal" (with discussion), *Statistical Science*, 9, 65–93.
- Gilks, W. R., and Wild, P. (1992), "Adaptive Rejection Sampling for Gibbs Sampling," *Applied Statistics*, 41, 337–348.
- Leonard, T., and Novick, M. R. (1986), "Bayesian Full Rank Marginalization for Two-Way Contingency Tables," *Journal of Educational Statistics*, 11, 33–56.
- MacGibbon, B., and Tomberlin, T. J. (1989), "Small Area Estimates of Proportions via Empirical Bayes Techniques," *Survey Methodology*, 15, 237–252.
- Maiti, T. (1997), "Hierarchical Bayes Estimation of Mortality Rates for Disease Mapping," Technical Report 546, University of Florida, Dept. of Statistics.
- Malec, D., Sedransk, J., and Tompkins, L. (1993), "Bayesian Predictive Inference for Small Areas for Binary Variables in the National Health Interview Survey," in *Case Studies in Bayesian Statistics*, eds. C. Gatsonis, J. S. Hodges, R. E. Kass, and N. D. Singpurwalla, New York: Springer-Verlag, pp. 377–389.
- McCullagh, P., and Nelder, J. A. (1989), *Generalized Linear Models* (2nd ed.), London: Chapman and Hall.
- Nandram, B., and Sedransk, J. (1993), "Bayesian Predictive Inference for a Finite Population Proportion: Two-Stage Cluster Sampling," *Journal of the Royal Statistical Society, Ser. B*, 55, 399–408.
- Prasad, N. G. N., and Rao, J. N. K. (1990), "The Estimation of the Mean Squared Error of Small-Area Estimators," *Journal of the American Statistical Association*, 85, 163–171.
- Spiegelhalter, D. J., Thomas, A., Best, N., and Gilks, W. R. (1995), "BUGS: Bayesian Inference Using Gibbs Sampling, Version 0.50," technical report, Cambridge University, Institute of Public Health, Medical Research Council Biostatistics Unit.
- Statistics Canada (1992), "The 1991 General Social Survey—Cycle 6: Health," in *Public Use Microdata File Documentation and User's Guide*, Ottawa: Statistics Canada.
- Stroud, T. W. F. (1987), "Bayes and Empirical Bayes Approaches to Small Area Estimation of Small Area Statistics," in *International Symposium on Small Area Statistics*, eds. R. Platek, J. N. K. Rao, C. E. Saerndal, and M. P. Singh, New York: Wiley, pp. 124–140.
- (1991), "Hierarchical Bayes Predictive Means and Variances With Application to Sample Survey Inference," *Communications in Statistics, Part A—Theory and Methods*, 20, 13–36.
- (1994), "Bayesian Inference From Categorical Survey Data," *Canadian Journal of Statistics*, 22, 33–45.
- Tierney, L., and Kadane, J. B. (1986), "Accurate Approximations for Posterior Moments and Marginal Densities," *Journal of the American Statistical Association*, 81, 82–86.
- Tsutakawa, R. K. (1985), "Estimation of Cancer Mortality Rates: A Bayesian Analysis of Small Frequencies," *Biometrics*, 41, 69–79.
- (1988), "Mixed Model for Analyzing Geographic Variability in Mortality Rates," *Journal of the American Statistical Association*, 83, 37–42.
- Tsutakawa, R. K., Shoop, G. L., and Marienfeld, C. J. (1985), "Empirical Bayes Estimation of Cancer Mortality Rates," *Statistics in Medicine*, 4, 201–212.
- Waller, L. A., Carlin, B. P., Xia, H., and Gelfand, A. E. (1997), "Hierarchical Spatio-Temporal Mapping of Disease Rates," *Journal of the American Statistical Association*, 92, 607–617.
- Zeger, S. L., and Karim, M. R. (1991), "Generalized Linear Models With Random Effects: A Gibbs Sampling Approach," *Journal of the American Statistical Association*, 86, 79–86.