Technical Report RM 679, Department of Statistics and Probability, Michigan State University

Experiences with Approximate Bayes Inference for the Poisson-CAR Model

Sarat C. Dass, Chae Young Lim and Tapabrata Maiti *
Department of Statistics & Probability
Michigan State University, East Lansing, MI 48824
Email: {sdass,lim,maiti}@stt.msu.edu

Running title: Approximate Bayes for the Poisson-CAR model

Abstract

In many areas of epidemiologic, demographic and geographical research, inference based on hierarchical spatial regression models is popular and important; for example, in disease mapping, environmental and health monitoring studies. Several estimation and inferential procedures have been proposed for these models, utilizing a variety of methods such as estimating equations, empirical Bayes and hierarchical Bayes. Hierarchical Bayes provides the full range of statistical inference (point as well as interval estimation) which may not be readily available in the other approaches. However, hierarchical Bayes is not problem-free and computations can be challenging in complex applications. Recently, an alternative method, namely the approximate Bayes, has been proposed to alleviate the problems with the hierarchical Bayes method. Approximate Bayes uses an integrated nested Laplace approximation to derive numerical approximations to various marginals of the full posterior distributions, thus avoiding Markov Chain Monte Carlo sampling completely. In this article, we compare and contrast between approximate Bayes, hierarchical Bayes and two other inferential methodologies in the context of hierarchical spatial regression models. Our emphasis is to investigate some of the claims made on approximate Bayes, namely the computational gain and the extent of automation, in the implementation. The differences have been demonstrated via simulation as well as through real examples.

Keywords: Approximate Bayes, Default Prior, Disease Mapping, Empirical Bayes, Estimating Equations, Hierarchical Bayes, Hierarchical Spatial Models.

1 Introduction

Mapping incidence and mortality from diseases such as cancer is an indispensable tool for epidemiologists to understand disease etiology. The standardized mortality ratio (SMR) is the primary

*Authors' names are in alphabetical order.

measure for displaying incidence maps. However, direct use of SMRs can be misleading since their estimates do not account for the varying population sizes over the region under consideration (Clayton and Kaldor, 1987). In their seminal paper, Clayton and Kaldor (1987) proposed shrinkage estimators for the SMR based on Poisson-gamma and Poisson-lognormal models. Thereafter, numerous extensions and variations of these basic models were proposed in the literature. We refer the reader to recent review articles and books by Lawson (2006), Elliot *et al.* (2000), Waller and Gotway (2004) and the special volume of *Statistics in Medicine* (2000) on this important issue.

The most important statistical problem (which is common to all models) is to estimate and make inference on the SMRs and model parameters. Two powerful approaches, namely the empirical Bayes (EB) and hierarchical Bayes (HB) methods, are quite popular in their own right. In EB, unknown model parameters are estimated from the observed data likelihood. However, EB provides approximate inference only and there is no (theoretically valid) method currently available for estimating EB confidence intervals in our context. The approach introduced in Clayton and Kaldor (1987) (CK) to estimate SMR is the EB method we investigate here.

On the other hand, the HB methodology is exact and accurate (in principle), and provides the fullest possibilities for inference (including confidence interval construction). There are however several difficulties in the HB inferential framework. HB requires the elicitation of a prior on the unknown model parameters, and therefore, can be sensitive to prior specification. Non-informative prior elicitation potentially provides a stable HB solution in absence of definite subjective information; for example, Sun *et al.* (1999) develops non-informative priors for CAR models. Another difficulty is that HB utilizes Markov Chain Monte Carlo (MCMC) techniques which can be computationally intensive to develop. In some cases, the MCMC chains can be slow to mix and the decision of whether the chain has converged can be sometimes dubious.

The main objective of this article is to search for viable alternatives which utilizes the theoretical correctness of HB but yet is computationally less intensive and user friendly. To this goal, we explored the approximate Bayes (AB) approach, a method based on *integrated nested Laplace approximation* recently proposed by Rue, Martino and Chopin (2009), (*Journal of Royal Statistical Society, Series B*, discussion paper). It is not possible to explore the impact of EB, HB and AB approaches on all available hierarchical spatial models in the literature. Instead, we investigate their impact on one of the most basic models which is popular and extremely useful, namely the Poisson-CAR model. The Poisson-CAR model is characterized by a spatial latent distribution for the expected SMRs, and conditional on the SMRs, the area specific counts follow a Poisson distribution.

Estimating equation based approaches (EE) avoid full distributional specifications by only requiring some moment assumptions, and hence, we also considered the EE approach here. An attractive property of the EE approach is that difficulties in maximizing the likelihood of the Poisson-spatial regression model are completely avoided. However, prediction of relative risks is difficult from this approach. In this paper, we consider the approach introduced in Yasui and Lele (1997) (YL) as one of EE methods.

The motivation of this work came from our current project of mapping lung cancer for all the counties in Michigan, USA, where we need a simple yet statistically feasible and valid procedure. Michigan is a highly segregated state by industrial zones. Thus, the disease map would help identify areas that need immediate government attention for various health issues. While working with the Michigan dataset, we also experimented with the Scottish lip cancer data (a data set that has been used heavily in the literature) using all the four methods EE, EB, HB and AB. The procedures are also compared using several simulation studies.

Based on the Poisson-CAR regression model, we find as expected that HB is sensitive to prior choice. Thus, we have considered *non-informative* priors for both HB and AB so that the inference is not sensitive with respect to the prior specification. Another advantage of using non-informative priors is that the Bayes estimates often have good frequentist properties. While implementing the Bayesian procedures via AB, it appears that each application (the model and the dataset) needs customization with a specific set of tuning parameters, contrary to HB. These settings of tuning parameters may significantly increase computational time depending on the inferential goals.

We organize the sections as follows. Section 2 describes the Poisson-CAR model briefly and mentions the inferential goals. Section 3 describes details of the four procedures EE, EB, HB and AB. Section 4 contains numerical results from the simulation studies as well as the real data sets. Finally, Section 5 presents the discussion followed by conclusions.

2 The Poisson-CAR Model

Consider n sites on a spatial domain. Let $\boldsymbol{Y} = (Y_1, \cdots, Y_n)^T$ and $\boldsymbol{E} = (E_1, \cdots, E_n)^T$ denote, respectively, the incidence counts and the expected population under risk for the n sites. We denote the i-th area-specific relative risk by θ_i, and $\boldsymbol{\theta} = (\theta_1, \cdots, \theta_n)^T$ to be the collection of all the area-specific relative risks. The *standardized mortality ratio* $SMR = \frac{Y}{E}$ is the maximum likelihood estimate (MLE) of $\boldsymbol{\theta}$ when Y_is are independent Poisson with mean $E_i\theta_i$ for $i = 1, 2, \cdots, n$.

It is well-known that the MLE is not a good measure of risk since it does not account for the differences in population sizes (which was highlighted in the seminal paper by Claton and Kaldor (1987)). Various smoothing models and their refinements have been proposed to overcome this drawback in the last twenty years or so. A typical property of smoothing models is to borrow strength from spatially related areas to come up with a more reliable estimate of the SMR. We consider here the basic model proposed by Clayton and Kaldor (1987) which has been used and modified by many other researchers in recent years.

Let $\boldsymbol{H}_i = (H_{i1}, \cdots, H_{ip})^T$ be a set of p covariates associated with site i, for $i = 1, 2, \cdots, n$, $\boldsymbol{\beta} = (\beta_1, \cdots, \beta_p)^T$ be the corresponding regression coefficients and $X_i = \log(\theta_i), i = 1, \cdots, n$ be the

log relative risks. The Poisson-CAR model is defined hierarchically as follows:

$$
\begin{aligned}
Y_i|X_i &\overset{ind}{\sim} Poisson(E_i \exp(X_i)) && (1)\\
X_i &= H_i^T\boldsymbol{\beta} + \epsilon_i && (2)\\
\boldsymbol{\epsilon} &\sim Normal\left(\mathbf{0}, \sigma^2\boldsymbol{D}(\gamma)\right); && (3)
\end{aligned}
$$

in (3), $\boldsymbol{\epsilon} = (\epsilon_1, \cdots, \epsilon_n)^T$ is taken to be distributed as the *conditional auto-regressive* (CAR) model (Besag (1974)) with covariance matrix $\boldsymbol{D}(\gamma) = (I - \gamma\boldsymbol{M}\boldsymbol{W})^{-1}\boldsymbol{M}$. Here, $\boldsymbol{W} = ((w_{ij}))$ is the $n \times n$ adjacency matrix that defines the neighboring (spatial) structure with $w_{ij} = 1$ if i and j are spatial neighbors, and 0, otherwise ($w_{ii} = 0$ by convention). The matrix $\boldsymbol{M} = ((m_{ij}))$ is a $n \times n$ diagonal matrix whose i-th diagonal entry is given by $\frac{1}{w_{i+}}$ where $w_{i+} = \sum_{j=1}^{n} w_{ij}$. The parameter $\gamma \in (-1, 1)$ defines the extent of spatial dependence and σ^2 is the measure of global variability. We refer the reader to Cressie and Chan (1989) and Yasui and Lele (1997) for detailed description and interpretation of the model and parameters.

In a typical disease mapping problem, the main interest is to infer the true relative risks θ_is, and the model parameters $\boldsymbol{\delta}^T = (\boldsymbol{\beta}^T, \gamma, \sigma^2)$. When mapping is of primary concern, the parameter of interest will be the θ_is. Instead, if we are interested in studying association between the exposure and outcome variables, or determine the spatial dependence, $\boldsymbol{\beta}$ and γ will be of primary interest. Confidence intervals is a general method of inference for θ_is and model parameters. In the subsequent paragraphs, we will compare several well-known methods with respect to inference in the hierarchical model of (1-3).

3 Methods of inference for the Poisson-CAR model

3.1 *Yasui and Lele (YL)*

Yasui and Lele (1997) developed an approach where the model parameters $\boldsymbol{\delta} = (\boldsymbol{\beta}^T, \gamma, \sigma^2)^T$ are estimated based on estimating equations. The estimating equation approach avoids difficulties involved with direct maximization of the likelihood (for example, the use of high dimensional integration) by defining moments rather than fully specifying the distribution. These difficulties are well recognized and documented in the literature (see, for example, Breslow and Clayton, 1993). We adopt YL for the Poisson-CAR model in (1-3) and compare it with the EB, HB and AB methods in the subsequent text.

The estimating equation for $\boldsymbol{\beta}$ is

$$\boldsymbol{H}^T\tilde{\boldsymbol{V}}^{-1}(\boldsymbol{Y}^* - \boldsymbol{H}\boldsymbol{\beta}), \tag{4}$$

where i-th element of $\boldsymbol{Y}^*$ is equal to $Y_i^* \equiv \log\{(Y_i + 1/2 - I_{\{Y_i=0\}}/4)/E_i\}$. Yasui and Lele (1997) determine $\boldsymbol{Y}^*$ as the bias-corrected version of log SMR. Also, in (4), $\tilde{\boldsymbol{V}}^{-1}$ is defined as

$$\tilde{\boldsymbol{V}}^{-1} = \begin{cases} E(var(\boldsymbol{Y}^*|\boldsymbol{\theta}))^{-1}\{E(var(\boldsymbol{Y}^*|\boldsymbol{\theta}))^{-1} + M^{-1}/\sigma^2\}^{-1}D^{-1}/\sigma^2, & \text{if } \sigma^2 > 0, \\ E(var(\boldsymbol{Y}^*|\boldsymbol{\theta}))^{-1} & \text{, if } \sigma^2 = 0, \end{cases} \tag{5}$$

where the i-th diagonal element $(var(\boldsymbol{Y}^*|\boldsymbol{\theta}))^{-1}$ is approximated by $E_i \exp(\boldsymbol{H}_i^T\boldsymbol{\beta} - \sigma^2 m_{ii}/2)$. If $\boldsymbol{\theta}$ is observed, the estimating equation for $\boldsymbol{\beta}$ (given γ) would be $\boldsymbol{H}^T D^{-1}(\boldsymbol{\theta}^* - \boldsymbol{H}\boldsymbol{\beta})$ where $\boldsymbol{\theta}^* = \log(\boldsymbol{\theta})$. Since $\boldsymbol{\theta}$ is not observable, it is replaced by the bias-corrected quantity $\boldsymbol{Y}^*$ and the corresponding optimal weight $\boldsymbol{H}^T(var(\boldsymbol{Y}^*))^{-1}$ is approximated by $\boldsymbol{H}^T\tilde{\boldsymbol{V}}^{-1}$. When $\boldsymbol{\theta}$ is known, the estimating equations for γ and σ^2, respectively, are

$$(\boldsymbol{\theta}^* - \boldsymbol{H}\boldsymbol{\beta})^T MW(I - \gamma MW)(\boldsymbol{\theta}^* - \boldsymbol{H}\boldsymbol{\beta})$$
$$(\boldsymbol{\theta}^* - \boldsymbol{H}\boldsymbol{\beta})^T(M^{-1} - \gamma W)(\boldsymbol{\theta}^* - \boldsymbol{H}\boldsymbol{\beta}) - n\sigma^2.$$

The above estimating equations are functions of $\boldsymbol{\theta}^*$ via $\theta_i^*, \theta_i^{*2}$ and $\theta_i^*\theta_j^*$ for $j \neq i$ and $i, j = 1, 2, \cdots, n$. Yasui and Lele (1997) replaced θ_i^*, θ_i^{*2} and $\theta_i^*\theta_j^*$ with Y_i^*, Y_i^{**} and $Y_i^*Y_j^*$, respectively, where Y_i^{**} is derived from the approximately unbiased estimate $\max\{0, (\log(Y_i+1/2))^2 - 1/(Y_i+1)\}$ of $(\theta_i^* + \log E_i)^2$.

The intention of Yasui and Lele (1997) was to propose an alternative method to the penalized and marginal quasi-likelihood (PQL and MQL) approaches of Clayton and Kaldor (1993). Yasui and Lele (1997) recognized that "the PQL and MQL approaches may require more computational time than the Gibbs sampler", and moreover, the theoretical properties of PQL and MQL were not well studied. Our take on the matter is that while the YL method is certainly computationally more efficient (since it does not require the inversion of large dimensional matrices), the estimating framework is not integrated in the sense that the model parameters are not estimated simultaneously. As a result, it is hard to assess how the uncertainty of estimating one parameter affects the others. Asymptotic properties of estimates in the YL framework are somewhat (but not completely) rigorous; for example, the asymptotic arguments hold for unbiased estimators of θ_i^* but no such exactly unbiased estimators are provided. What YL proposed in applications is to use approximately unbiased estimators for which the asymptotic results may or may not hold. Further, properties of associated confidence intervals are only given for $\boldsymbol{\beta}$, but not for the other parameters (for example, for γ and σ^2).

3.2 *Clayton and Kaldor (CK)*

Clayton and Kaldor (1987) proposed an EB-type approach where the estimator of relative risk is the conditional expectation of θ_i given $\boldsymbol{Y}$. The model parameters are estimated using the EM-algorithm. The conditional expectation of θ_i given the data is derived in an approximated closed form and the estimated model parameters are plugged-in for the corresponding unknown quantities in the expression. The posterior distribution of $\boldsymbol{X} = (X_1, X_2, \cdots, X_n)$ (see (2)) conditional on the data is approximated by a multivariate normal density with mean $\boldsymbol{\mu}_X$ and covariance matrix V_X; the expressions for the mean and covariance are

$$\begin{aligned} \boldsymbol{\mu}_X &= V_X\left(\frac{1}{\sigma^2}D^{-1}\boldsymbol{H}\boldsymbol{\beta} - \psi''(\tilde{\boldsymbol{X}})\tilde{\boldsymbol{X}} + \psi'(\tilde{\boldsymbol{X}})\right) \quad \text{and} \\ V_X &= \left(\frac{1}{\sigma^2}D^{-1} - \psi''(\tilde{\boldsymbol{X}})\right)^{-1}, \end{aligned}$$

where $\tilde{\boldsymbol{X}} = \log\left(\frac{Y_i+1/2}{E_i}\right)$, $\psi^{'}(\tilde{\boldsymbol{X}}) = -1/2$, and $\psi^{''}(\tilde{\boldsymbol{X}}) = -(Y_i + 1/2)$ for the Poisson likelihood. The expectation step of the EM algorithm has also been similarly approximated in their paper. We refer the reader to their paper for more details. The main advantage of the CK approach is that the point estimates of $\boldsymbol{\delta}$ and θ_i, $i = 1, 2, \cdots, n$ are easily obtained. However, measures of uncertainty of the point estimates as well as methods for construction of confidence intervals are not available. Some efforts have been made in this regard, but to our knowledge, nothing is satisfactory. For example, Hall and Maiti's (2006) parametric bootstrap approach could be used but this needs further theoretical investigation to justify their approach. No effort has been made as well for the improved inference regarding regression parameters in the context of spatial regression.

3.3 *Hierarchical Bayes* (HB)

Unlike the CK approach, complete inference is possible in the HB framework based on the posterior distribution of unknown quantities provided one elicits a suitable prior for the model parameters. However, we found that the inference can be sensitive to the choice of prior elicitation in this Poisson-CAR model. In fact, Gelman (2006) argued against the use of standard vague (weakly informative) prior in a simple variance component model which is a special case of the Poisson-CAR model. This motivated us to propose a default prior on $\boldsymbol{\delta}$ which combines non-informative (e.g., flat) priors on $\boldsymbol{\beta}$ and γ with a default prior on σ^2. Following a Taylor's expansion in Dass et. al (2009) around a $\boldsymbol{X}^* = (X_1^*, X_2^*, \cdots, X_n^*)^T$, one can have an explicit expression for the conditional mean of X_i for the model in (1-3) which is a weighted combination of data and prior contributions. The approximated posterior mean of X_i has the expression

$$\frac{K_i}{K_i + 1/\sigma^2}\left(Y_i - h_i^{'}(X_i^*) + X_i^* h_i^{''}(X_i^*)\right) + \frac{1}{K_i + 1/\sigma^2} X_i^{CAR} \tag{6}$$

where $K_i = h_i^{''}(X_i^*)/w_{i+}$, $h_i(x) = E_i \exp(x)$ is the inverse of the link function for the Poisson density at the i-th site, and X_i^{CAR} is the prior conditional mean of X_i given the rest of X_j, $j \neq i$. Since K_i varies across all the sites, we replace the average $\omega_0 = \frac{1}{n}\sum_{i=1}^{n} K_i$ for each K_i in expression (6). Requiring the weights to have a uniform prior leads to the following default prior elicitation on σ^2:

$$\pi_D(\sigma^2) \propto \frac{1}{(1 + \omega_0\sigma^2)^2}. \tag{7}$$

Note that $\pi_D(\sigma^2)$ is a proper density. It can be shown that posterior distribution is proper based on the above prior specification of $\boldsymbol{\delta}$. Incidentally, the above prior is similar to half-t prior introduced in Gelman (2006) for the random effects model. We compare the parameter estimates based on different prior choices using simulated data and results is shown in Section 4.1. Samples from posterior of $\boldsymbol{\delta}$ and the relative risk $\boldsymbol{\theta}$ are obtained using Markov Chain Monte Carlo (MCMC) techniques and subsequently used for inference. We implemented the MCMC algorithm in MATLAB. Note that we could use BUGS which is well established program for the MCMC method and the

CAR model is available within the program. However, we need to use 'zeros' tricks for a general form of the prior distribution which is described in the User Manual. As mentioned in the Manual, this method produces high autocorrelation and poor convergence and so, we did not pursue with BUGS in this study.

3.4 *Approximate Bayes* (AB)

Recently Rue et al. (2009) advocated the use of approximate Bayes methods as an alternative to HB for posterior based inference. They claim that one of the benefits of AB is " ... computational: where Markov Chain Monte Carlo need hours or days to run, our approximations provide more precise estimates in seconds or minutes". Our objective here is to compare and contrast AB with HB in terms of statistical accuracy and the computational time involved when obtaining posterior measures of inference (e.g., confidence intervals) for the Poisson-CAR model. It will be demonstrated in the next section that the picture is different than what is anticipated in this special case of spatial regression set-up.

The AB approach has three important approximation steps which we describe now. The posterior of $\boldsymbol{\delta}$ given $\boldsymbol{Y}$ for the Poisson-CAR model is

$$\pi(\boldsymbol{\delta}|\boldsymbol{Y}) = \frac{\pi(\boldsymbol{Y}|\boldsymbol{X})\pi(\boldsymbol{X}|\boldsymbol{\delta})\pi(\boldsymbol{\delta})}{\pi(\boldsymbol{X}|\boldsymbol{Y},\boldsymbol{\delta})}, \tag{8}$$

which holds for any value of $\boldsymbol{X}$. The first step of AB is to approximate the posterior distribution of log relative risks, $\pi(\boldsymbol{X}\,|\,\boldsymbol{Y},\,\boldsymbol{\delta})$, by a multivariate Gaussian density $\pi_{GG}(\boldsymbol{X}\,|\,\boldsymbol{Y},\,\boldsymbol{\delta})$. The posterior density of $\boldsymbol{\delta}$ is then approximated by

$$\widehat{\pi}(\boldsymbol{\delta}|\boldsymbol{Y}) = \left.\frac{\pi(\boldsymbol{Y}|\boldsymbol{X})\pi(\boldsymbol{X}|\boldsymbol{\delta})\pi(\boldsymbol{\delta})}{\pi_{GG}(\boldsymbol{X}|\boldsymbol{Y},\boldsymbol{\delta})}\right|_{\boldsymbol{X}=\boldsymbol{X}^*(\boldsymbol{Y},\boldsymbol{\delta})}, \tag{9}$$

where $\boldsymbol{X}^*(\boldsymbol{Y},\boldsymbol{\delta})$ is the mode of $\pi_{GG}(\boldsymbol{X}\,|\,\boldsymbol{Y},\boldsymbol{\delta})$ for a given $\boldsymbol{\delta}$. Finally, the marginal posterior distribution of the log relative risks is approximated using numerical integration of an integrand which is a product of the approximations in the first two steps. The key idea of AB is to replace the denominator of (8), $\pi(\boldsymbol{X}|\boldsymbol{Y},\boldsymbol{\delta})$, with a Gaussian approximation $\pi_{GG}(\boldsymbol{X}|\boldsymbol{Y},\boldsymbol{\delta})$ and evaluate it at the mode $\boldsymbol{X}^*(\boldsymbol{Y},\boldsymbol{\delta})$. Since no exact closed form is available for $\boldsymbol{X}^*(\boldsymbol{Y},\boldsymbol{\delta})$, Rue et al (2009) computes this mode using the Newton-Raphson algorithm. Note that the multivariate Gaussian approximation $\pi_{GG}(\boldsymbol{X}\,|\,\boldsymbol{\delta},\boldsymbol{Y}\,)$ on $\boldsymbol{X}$ forces each of the marginal distributions of X_i to be normal which we denote by $\pi_{GG,i}(X_i\,|\,\boldsymbol{\delta},\boldsymbol{Y})$ for $i = 1, 2, \cdots, n$. The marginal posterior densities of X_i and $\boldsymbol{\delta}$ are obtained by integrating out the irrelevant terms in the full posterior. An approximate expression for the marginal is

$$\pi(X_i|\boldsymbol{Y}) \;\; = \;\; \int \pi(X_i|\boldsymbol{Y},\boldsymbol{\delta})\pi(\boldsymbol{\delta}|\boldsymbol{Y})d\boldsymbol{\delta} \approx \sum_k \pi_{GG,i}(X_i\,|\,\boldsymbol{\delta}_k,\,\boldsymbol{Y}\,)\widehat{\pi}(\boldsymbol{\delta}_k\,|\,\boldsymbol{Y})\Delta_k \tag{10}$$

which is evaluated using numerical integration on a set of $\boldsymbol{\delta}$-grid points, $\boldsymbol{\delta}_k$, with area weights Δ_k for $k = 1, 2\cdots, K$. A similar numerical integration technique is used for the evaluation of the

marginal

$$\pi(\delta_i \,|\, \boldsymbol{Y}) \;\; = \;\; \int \pi(\boldsymbol{\delta}|\boldsymbol{Y}) d\boldsymbol{\delta}_{-i} \approx \int \widehat{\pi}(\boldsymbol{\delta}|\boldsymbol{Y}) d\boldsymbol{\delta}_{-i}. \tag{11}$$

Rue et al (2009) proposed other methods such as Laplace and simplified Laplace approximations to improve the accuracy when evaluating $\pi(X_i|\boldsymbol{Y})$. Eidsvik et al. (2009) developed AB specifically for spatial generalized linear mixed models. Our study is in the same spirit of Eidsvik et al. (2009) although our model and inferential goals are somewhat different.

The accuracy of the above numerical integration steps requires a good choice of $\boldsymbol{\delta}$ evaluation points. The approach suggested in Rue et al. (2009) is

- **STEP 1:** locate the mode of $\widehat{\pi}(\boldsymbol{\delta}|\boldsymbol{Y})$, $\boldsymbol{\delta}^*$ say, by optimizing $\log \widehat{\pi}(\boldsymbol{\delta}|\boldsymbol{Y})$ with respect to $\boldsymbol{\delta}$,
- **STEP 2:** compute the negative Hessian matrix $\boldsymbol{S}$ at $\boldsymbol{\delta} = \boldsymbol{\delta}^*$ and consider the spectral value decomposition of $\boldsymbol{S}^{-1} = Q\Lambda Q^T$. The centered and scaled variable $\boldsymbol{z}$ is defined as

$$\boldsymbol{z} = Q^T \Lambda^{-1/2} (\boldsymbol{\delta} - \boldsymbol{\delta}^*) \qquad \text{or} \qquad \boldsymbol{\delta}(\boldsymbol{z}) = \boldsymbol{\delta}^* + Q\Lambda^{1/2}\boldsymbol{z}. \tag{12}$$

- **STEP 3:** find a collection of $\boldsymbol{z}$, $\mathcal{Z}$, such that the corresponding $\boldsymbol{\delta}(\boldsymbol{z})$ points are located around the mode $\boldsymbol{\delta} = \boldsymbol{\delta}^*$. Starting from $\boldsymbol{z} = 0$ ($\boldsymbol{\delta} = \boldsymbol{\delta}^*$), each component entry of $\boldsymbol{z}$ is searched in the positive and negative directions in step sizes of η_z. All z-points satisfying

$$\log \widehat{\pi}(\boldsymbol{\delta}(\boldsymbol{0})|\boldsymbol{Y}) - \log \widehat{\pi}(\boldsymbol{\delta}(\boldsymbol{z})|\boldsymbol{Y}) < \eta_\pi$$

  are taken to be in $\mathcal{Z}$.

- **STEP 4:** evaluate $\pi(\delta_i \,|\, \boldsymbol{Y})$ for a fine grid of δ_i points based on values in $\mathcal{Z}$. This last step is needed for any inference procedure requiring the evaluation of the posterior density values, such as obtaining HPD confidence sets.

Note that the values of η_z and η_π should be appropriately tuned to produce accurate approximations. The trade-off here is numerical accuracy versus computational time. The above standardization technique allows for the evaluation of $\pi(\boldsymbol{\delta} \,|\, \boldsymbol{Y})$ at the points $\boldsymbol{\delta} = \boldsymbol{\delta}(\boldsymbol{z})$ for $\boldsymbol{z} \in \mathcal{Z}$ only. The map $\boldsymbol{\delta}(\mathcal{Z}) \equiv \{\boldsymbol{\delta}(\boldsymbol{z}),\, \boldsymbol{z} \in \mathcal{Z}\}$ typically forms an ellipsoidal region in the $\boldsymbol{\delta}$-space. To compute $\pi(\delta_i \,|\, \boldsymbol{Y})$, the integral in (11) has to be evaluated for each fixed point $\delta_i = \delta_0$. Consequently, this method requires a substantial number of points in $\boldsymbol{\delta}(\mathcal{Z})$ having δ_0 as the value of its i-th component. This is almost always not the case, thus necessitating the development of an interpolation scheme to bypass this difficulty.

AB constrasted with HB: Once the model and the prior distributions are specified, unlike the HB, the approximate Bayes approach needs to have following tuning parameters pre-specified: The choice of z-values, the choice of η_z and η_π, and the preference of interpolation as stated in the previous paragraph. This is additional specification on top of HB specifications like the choice of prior.

4 Numerical Properties

The numerical studies are grouped by three different aspects of the Poisson-CAR model: (1) elicitation of priors and selection of a suitable robust prior, (2) comparison between different methods of inference, and (3) the real data applications. Several simulation studies were conducted to compare the performance of the proposed non-informative prior with other subjective priors, and to compare parameter estimates and their confidence/credible intervals obtained from their respective posteriors. Our benchmark approach is the well-established MCMC based HB. For HB, we checked all convergence of the chains up to satisfactory levels.

4.1 Prior Elicitation

Following Yasui and Lele (1997), we generated $n = 10 \times 10$ grid points. The adjacency matrix $\boldsymbol{W}$ is defined in terms of the four nearest neighbors corresponding to each site of interest. The regression covariates are $H_i = (1, h_i)^T$ where h_i is generated from a normal distribution with mean zero and variance 0.5. The regression coefficients are taken as $\boldsymbol{\beta} = (\beta_0, \beta_1)^T = (0.1, 0.3)^T$, the spatial dependence parameter $\gamma = 0.95$ and variance $\sigma^2 = 0.2$. The expected counts E_i are generated from uniform on $[10, 30]$, and given E_i, Y_i is generated from the Poisson distribution with mean $E_i\theta_i$. We generate 200 replications from this model specification.

To assess the performance with respect to prior choice, several choices of subjective priors were taken for σ^2 in addition to the default prior, π_D, and the half-Cauchy prior of Gelman (2006). The subjective choice was the inverse gamma prior $IG(a, b)$ with several different choices of a and b; here, we define the IG prior as $1/IG(a, b) \sim G(a, b)$ where $G(a, b)$ has mean ab and variance ab^2. We consider five sets of a and b: $(a, b) = (0.01, 0.01)$, $(0.1, 0.1)$, $(0.1, 100)$, $(0.01, 500)$ and $(0.01, 10)$. The choices of a and b were selected to represent prior information ranging from weakly to highly informative. To represent weakly informative priors, one might choose small and large values of a and b, respectively, which are the last three choices above. The first two choices with small a and b has been considered in several works, for example, Natarajan and Kass (2000). We found that the results of inference is very sensitive to the prior specification. The first two combinations of (a, b) gave greatly varying results while the last three gave fairly similar results. For illustration and to save space, we selected the first two combinations and $(a, b) = (0.01, 500)$ for reporting purposes. The half-Cauchy prior of Gelman (2006) is

$$\pi_C(\sigma^2) \propto \frac{1}{\sigma(1 + (\sigma/A)^2)}.$$

For the half-Cauchy prior, we set $A = 100$ which makes it weakly informative. We tried several values of A ranging from $A = 0.1$ to $A = 100$ but found that the inference is fairly robust. Thus, we report only results based on one A specification, namely, $A = 100$. Note that the prior π_D is fully automatic with no hyper-parameters to be specified. In each replication, we ran 3 Gibbs samplers for $10,000$ iterations and computed results using the last $3 \times 2,000 = 6,000$ samples. Convergence has been checked by monitoring the Gelman-Rubin's R^2 statistic. Table 1 gives several

statistical measures of performance, namely, the bias, mean-squared error (MSE) and coverage for 90% highest posterior density (HPD) sets. The average of the maximum-a-posteriori estimates (MAPs) computed from the 200 replications is also reported.

The SMR column in Table 1 reports averages over all $n = 100$ sites on the spatial grid. Clearly, the two informative prior choices (small a and b) on σ^2 behaves erratically. For example, although the coverage for $\boldsymbol{\beta}$ is reasonable, the coverages for γ and σ^2 are meaningless. This could be due to the high bias and MSE values for these two parameters. The three weakly informative prior choices give fairly sensible results and are comparable to the half-Cauchy (with large A) and default priors. The overall conclusion from Table 1 is that the weakly informative IG priors as well as the half-Cauchy and default priors are sensible priors to use as they give good and consistent results. In the subsequent numerical simulations, we use the default prior as the prior of choice for the HB and AB approaches.

[Table 1 approximately goes here]

4.2 A Comparison Between Different Methods

After selecting the prior, our objective is to implement and compare the HB and AB methodologies with respect to model parameter inference. In the Poisson-CAR regression model, inference on the regression parameters are usually of primary interest (rather than prediction). Since the CK and YL approaches are viable methods for estimation and inference on regression parameters, we also compare their performance with HB and AB. Table 2 reports the results. The YL method provides estimates, standard errors and confidence intervals for $\boldsymbol{\beta}$ based on asymptotic arguments as well as estimates for γ and σ^2. However, no confidence intervals are provided for σ^2 and γ. The CK method provides estimates of model parameters and relative risks but confidence intervals for them are unavailable. In contrast, the HB and AB methods provide estimates, standard errors and credible intervals for all model parameters as well as estimates of the relative risks, $\boldsymbol{\theta}$. This is possible since inference is carried out using samples from the posterior density in the HB method. In the AB method, numerical approximations to various marginal posterior densities are available. Since one of the main purposes is to compare AB and HB in terms of implementation, we have taken two versions of AB depending on their tuning parameters. The first and second versions of AB, AB1 and AB2, respectively, correspond to the specifications (1) $\eta_z = 1$, $\eta_\pi = 5$ and (2) $\eta_z = 0.5$, $\eta_\pi = 20$. The specifications of AB2 lead to a finer grid for numerical evaluations which result in higher accuracy but longer computational time. In our simulation study, the computational times are as follows. On a computer with Pentium IV dual processor, the approximate running time for the MCMC in HB is 25 min per replication. For AB1, the total time for one replication is 34 minutes whereas for AB2, it is 3.83 hours. The first three steps (i.e., **STEPS 1-3**) in AB1 take only 43 seconds but to complete **STEP 4** (needed for full inference), the time is 34 minutes. Similarly, for AB2, the times are 3.28 and 3.83 hours, respectively. The computational time depends on the size of $\mathcal{Z}$ which may vary from replication to replication in the simulation study.

[Table 2 approximately goes here]

Estimation of $\boldsymbol{\beta} = (\beta_0, \beta_1)^T$ in terms of the bias and mean square error (MSE) is best for CK followed by AB2 and HB, with AB1 giving comparable results to HB. For the estimation of γ, YL performs the best followed by HB, $AB2$ and CK in terms the MSE. For σ^2, CK is the best in terms of MSE followed by HB, AB and YL. For estimating the coverage, YL performed equally with HB and AB2 for β_1 but poorly for β_0. The coverage of HPD sets for all model parameters for AB1 are way off compared to the nominal coverage 90%. The coverage under AB2 is fairly comparable with HB for β_1 and σ^2, but still way off for β_0 and γ.

The above observations indicate that there is a considerable difference between AB and HB particularly for construction of HPD sets. The discrepancies can be reduced at the expense of increased amount of computational steps and time. To our surprise, we did not experience the striking computational advantage of the AB method as mentioned by Rue et al (2009) in our simulation studies. In searching for the reason behind this surprising behavior, we find this is due to accurate estimation of the tail area of the posterior distributions for computing the HPD sets. Figure 1 provides density plots of a randomly selected replicate (out of the 200) for the marginal posterior distributions of all model parameters under AB1 and AB2. On each plot, samples from the posterior in the HB method was used to compute a kernel density estimate and this was overlaid in each panel. While the plots for β_1 nicely coincide, discrepancies are clear for β_0, γ and σ^2 where the difference in coverage are observed. Note that the point estimates are fairly comparable with HB even under AB1 because the center of the posterior distributions are well matched. It is in the tails that the difference is the largest, and the tail regions are crucial for the construction of the HPD sets. This explains the differences in inference. Lawson, in the discussion of Rue et al (2009), queried this potential drawback of the AB method which we explicitly establish here.

[Figure 1 goes here]

If one is interested only in point estimates of the model parameters, fine tuning of η_z and η_π is not needed. The posterior distribution can be evaluated only on a small number of points in $\mathcal{Z}$ to yield comparable estimates in a relatively short time. For example, to get an estimate of β_0, one can apply AB1 based on steps (1-3) on the transformed grid and compute

$$E(\beta_0 \,|\, \boldsymbol{Y}) = \int_{\boldsymbol{\delta}} \beta_0 \, \pi(\boldsymbol{\delta} \,|\, \boldsymbol{Y}) \, d\boldsymbol{\delta} \approx \int_{\mathcal{Z}} \beta_0(\boldsymbol{z}) \, \widehat{\pi}(\boldsymbol{\delta}(\boldsymbol{z}) \,|\, \boldsymbol{Y}) \, d\boldsymbol{\delta}(\boldsymbol{z})$$

in less than a minute (approximately 43 seconds). The time reduction is from the fact that **STEP 4** is avoided completely. However, for construction of HPD sets, marginal posterior densities need to be evaluated for a number of β_0 points on a grid which necessitates **STEP 4** and the interpolation scheme mentioned after **STEP 4** earlier to transform $\mathcal{Z}$ back to the rectangular grid.

4.3 Real example: Scottish Lip Cancer data

In this subsection, we like to report our experience with two real examples using all the procedures discussed previously. Our first example is the Scottish lip cancer dataset which has been analyzed by many leading researchers including Clayton and Kaldor (CK) (1987) and Yasui and Lele (YL) (1997).

The Scottish lip cancer incidence data consists of the observed and expected cases of lip cancer during the 6 years from 1975 to 1980 in each of 56 counties of Scotland along with the percentage of the work force employed in agriculture, fishing, or forestry (AFF). The Poisson-CAR model is considered with one covariate $h_i = \text{AFF}/10$ whose information was not incorporated by CK. While HB and AB2 are in close agreement except for the HPD set of β_0. Note that AB1 has larger discrepancy compared to AB2. The computational times are as follows: The total time taken for HB is 22 minutes; for AB1 and AB2, the total time is 37 minutes and 5.16 hours, respectively. These results are consistent with our simulation study in Section 4.1 and are reported in Table 3; note that the numbers reported under YL are adopted from their original paper. Interestingly, YL reports some results in their paper that are not in close agreement with HB and AB. For example, the estimate of β_1 is 0.08 which is significantly different from the HB and AB approaches (both report the value at around 0.4). The 90% confidence interval for β_1 is $(0.08 \pm 1.65 \times 0.23)$, concluding that β_1 is insignificant. However, both HB and AB results show that the covariate information is strongly significant at 90% level. The estimates of β_0 and σ^2 for YL are also significantly different from HB and AB.

[Table 3 approximately goes here]

4.4 Michigan Lung Cancer data

We give another real data example to illustrate the difficulty and sensitivity involved in selection of the tuning parameters for AB. Our main aim of investigation is to find out whether the discrepancy between HB and AB remains the same from application to application given a fixed set of tuning parameters (for AB). Note that both HB and AB use the same prior elicitation here, namely the default prior, so the differences, if any, will be due to tuning parameter (not prior) specifications. The dataset we consider is lung cancer mortality incidences in 83 Michigan counties from 2001 to 2005, which were obtained from the SEER database (*URL: seer.cancer.gov*). Cancer mortality incidence is rare enough relative to the population in each county so that a Poisson-CAR model is reasonable. The expected count is computed by taking each county's age distribution into account, which is available from U.S. Census 2000. The expected age-adjusted number of deaths due to lung cancer in each county is calculated according to the formula in Jin et al (2005).

We consider one covariate at the county level, namely, the proportion of the population under poverty. This information is available from the SEER site at the county level based on Census 2000. Better covariate information (such as smoking status) for predicting lung cancer is possible but this

is currently unavailable at the county level. Another argument for choosing poverty is that a simple correlation between poverty rate and SMR due to lung cancer is reasonably high at 0.40. Thus, for the purpose of illustrating the issues with AB, we consider poverty as a covariate. Figure 2 shows the spatial distributions of observed SMRs and poverty over different Michigan counties which can be seen to have similar spatial distributions. Running the analyses with the four approaches, Table 4 shows that the coefficient of poverty is significant for AB, HB and YL.

[Figure 2 approximately goes here]

[Table 4 approximately goes here]

It is natural to take the HB approach as the benchmark. We have checked convergence of the MCMC samplers until they reached satisfactory levels. We applied the tuning parameter specifications of AB as set in the Scottish lip cancer dataset and the simulation results. However, it turns out even AB2 has severe disagreements with HB in this case. Since σ^2 is smaller in this example, this is perhaps the reason for such a discrepancy; the $\boldsymbol{\delta}$-grid for AB is too coarse to give rise to accurate numerical evaluations. Based on the estimated value of σ^2, we refined the grid by specifying $\eta_z = (0.5, 0.5, 0.25, 0.125)$, thus increasing the resolution of the $\mathcal{Z}$ grid. We call this AB3. The refinement reduced differences with HB but increased computational time significantly. The computational times are 19 minutes for HB, 22 minutes for AB1, 46.4 minutes for AB2, and 6.30 hours for AB3. This example clearly illustrates that besides the computational differences, the tuning parameters need to be chosen carefully in AB. In contrast, HB is more automatic once the prior is specified. The smoothed estimates of SMR are displayed in Figure 3. Apparently, Figure 3 does not convey any noticeable difference because they are generated based on point estimates where there is not much difference between HB and AB. However, the difference between the two methods will be prominent in terms of mapping HPD sets. In HB, obtaining HPD sets for all 83 counties is definitely possible since we have posterior samples from all the counties. However, this is not computationally feasible in AB since one can well imagine the computing time required (from previous discussions) for all 83 Michigan counties. This naturally prohibits any statistical comparisons between counties, for example the determination of high risk zones, via AB.

[Figure 3 approximately goes here]

5 Discussion and Conclusion

In this article, we have compared the EE, EB, HB and AB approaches in terms of their ability to make inference for the Poisson-CAR model. We have outlined the current status of all the four procedures in this context and have highlighted the pros and cons associated with each of them. Clearly, when only point estimates of the model parameters are of interest, EE is the simplest to implement. However, as one proceeds to require further statistical inference (for example, calculating standard errors and confidence intervals), one needs to look at HB and AB as alternative

procedures. Our main objective was to find a procedure which would be simple, flexible and easy to implement. Until recently, the Gibbs sampling based HB was proven to be the most powerful method in this setup. However, the HB is not without difficulties. For example, even after selecting the appropriate prior, the MCMC implementation and check for convergence can be challenging although there are fairly reasonable guidelines available in the literature. Thus, for searching an alternative and computationally less demanding procedure, we were led to the recently proposed AB. However, we faced several obstacles in its implementation.

To our surprise, the AB procedure took a lot more time than the well established HB approach at least in this setup. For the simple Poisson-CAR model, the time difference was more than several hours depending on the degree of inferential accuracy one is aiming for. This problem is further magnified when one is required to obtain two or more HPD intervals, for example in the case of the 83 Michigan counties. Although AB does not have checks on convergence, it requires the user to define several tuning parameters in its implementation. There is no clear guideline to selecting these tuning parameters. While some intuitive specifications of tuning parameters can be set for simulations (since the parameter values are known), this is more difficult in the case of real data and can throw off inference results if they are not reasonably set.

The AB approach has been shown to be numerically close to point estimates based on marginal posterior distributions but there is no theory nor numerical studies that indicate that the AB approximation is close to the true full posterior distribution. We have demonstrated in this paper that even for marginal posterior distributions, the tails do not match easily. Although we believe that it is possible to reach as close as one desires at the tails, it is not clear how to prepare a guideline for this (for example, selecting the right resolution for the $\boldsymbol{\delta}$-grid). Further, error bounds are not available from previous works on AB. Thus, we face a major hurdle in the AB implementation.

Our conclusion is that AB is useful for deriving point estimates with potential time savings compared to HB but for full inference (such as constructing of HPD sets), the implementation of AB is both time consuming and possibly inaccurate. It may happen that there could be an efficient way of coding using lower level languages such as $C++$ which could eventually reduce the total computational time. However, HB could have the same advantage if it were written in $C++$. In our experiments, we used MATLAB which is a commonly used platform for statistical computations.

References

BESAG, J.E. (1974). Spatial interaction and the statistical analysis of latice systems (with discussion). *J. Roy. Statist. Soc. Ser. B* **36** 192-236.

BRESLOW, N.E., and CLAYTON, D.G. (1993). Approximate inference in generalized linear mixed models. *J. Amer. Statist. Assoc.* **88** 9-24.

CRESSIE, N. and CHAN, N.H. (1989). Spatial modeling of regional variables. *J. Amer. Statist. Assoc.* **84** 393401.

CLAYTON, D. and KALDOR, J. (1987). Empirical Bayes estimates of age-standardized relative risks for use in disease mapping. *Biometrics* **43** 671-681.

DASS, S., LIM, C-Y., and MAITI, T. (2009). Default Bayesian analysis for hierarchical spatial multivariate generalized linear mixed models. Tech Rep. Dept. of Stat. and Prob., Michigan State University.

EIDSVIK, J., MARTINO, S., and RUE, H. (2009). Approximate Bayesian inference in spatial generalized linear mixed models. *Scan. J. Statist.* **36**, 1-22.

ELLIOT, P., WAKEFIELD, J.C., BEST, N.G., and BRIGGS, D.J. (Eds.) (2000). *Spatial Epidemiology: Methods and Applications*. London: Oxford University Press.

GELMAN, A. (2006). Prior distributions for variance components in hierarchical models. *Bayesian Statist.*

JIN, X., CARLIN, B.P., AND BANERJEE, S. (2005). Generalized hierarchical multivariate CAR models for areal data. *Biometrics* **61** 950-961.

LAWSON, A. (2006). *Statistical Methods in Spatial Epidemiology* (2nd ed.). New York: Wiley.

NATARAJAN, R. and KASS, R. E. (2000). Reference Bayesian methods for generalized linear mixed models. *J. Amer. Statist. Assoc.* **95** 227-237.

RUE, H., MARTINO, S. and CHOPIN, N. (2009). Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations (with discussion). *J. R. Statist. Soc. Ser. B* **71**, 319-392.

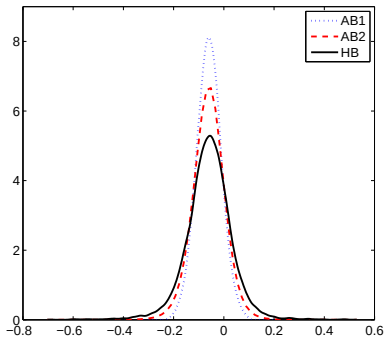
SUN, D., TSUTAKAWA, R.K., and SPECKMAN, P.L. (1999). Bayesian inference for CAR(1) models with noninformative priors. *Biometrika* **86** 341-350.

WAKEFIELD, J. (2007). Disease mapping and spatial regression with count data. *Biostatistics* **8** 158-183.

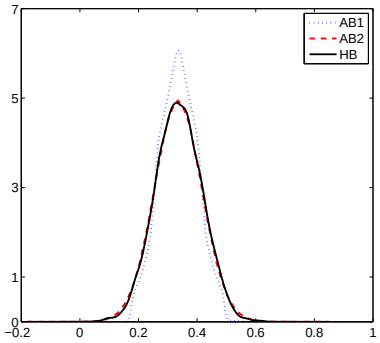
WALLER, L.A., and GOTWAY, C. (2004). *Applied Spatial Statistics for Public Health Data.* New York: Wiley.

YASUI, Y., and LELE, S. (1997). A regression method for spatial disease rates: An estimating function approach. *J. Amer. Statist. Assoc.* **92**, 21-32.

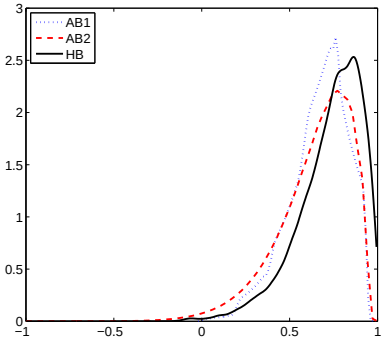
Figure 1: Posterior Densities from the AB and HB method



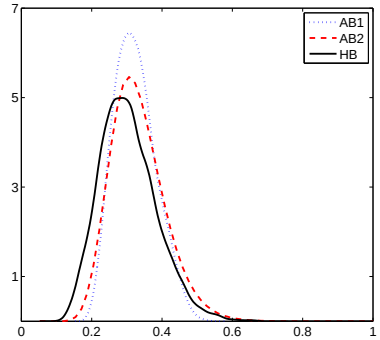
(a) β_0



(b) β_1



(c) γ



(d) σ^2

Table 1: Comparisons of different choices of prior specifications.

Prior Specification		Parameter (True Value)				SMR
		$\beta_0(0.1)$	$\beta_1(0.3)$	γ (0.95)	σ^2 (0.2)	$\bar{\theta}$
	Bias$\times 10^3$	-87.03	39.40	-783.49	2527.24	-1.47
$IG(a,b)$	MSE$\times 10^3$	16.84	6.79	619.66	6392.90	54.81
$a = 0.01$, $b = 0.01$	Average of MAPs	0.02	0.32	0.18	2.62	1.07
	Coverage	92.5%	100%	0%	0%	89.5%
	Average width of HPD set	0.36	0.60	0.84	1.30	0.75
	Bias$\times 10^3$	-65.07	31.84	-496.76	407.16	-1.94
$IG(a,b)$	MSE$\times 10^3$	13.23	5.42	280.52	167.78	41.48
$a = 0.1$, $b = 0.1$	Average of MAPs	0.04	0.32	0.52	0.58	1.07
	Coverage	67.5%	96.5%	20.5%	0%	91.8%
	Average width of HPD set	0.27	0.32	0.76	0.34	0.69
	Bias$\times 10^3$	-45.10	27.90	-197.52	4.21	-2.91
$IG(a,b)$	MSE$\times 10^3$	10.84	4.88	62.78	2.73	35.29
$a = 0.01$, $b = 500$	Average of MAPs	0.06	0.32	0.87	0.19	1.08
	Coverage	84.0%	89.0%	94.0%	92.0%	89.1%
	Average width of HPD set	0.31	0.22	0.49	0.19	0.59
	Bias$\times 10^3$	-47.05	27.32	-209.41	13.53	-3.06
Half Cauchy	MSE$\times 10^3$	11.06	4.84	69.14	3.00	35.32
$A = 100$	Average of MAPs	0.06	0.32	0.85	0.19	1.08
	Coverage	82.5 %	89.5 %	92.0 %	92.5 %	89.3%
	Average width of HPD set	0.30	0.22	0.50	0.20	0.59
	Bias$\times 10^3$	-46.20	27.25	-200.42	4.57	-3.10
Default	MSE$\times 10^3$	10.99	4.83	63.51	2.54	35.27
	Average of MAPs	0.06	0.32	0.86	0.18	1.08
	Coverage	84.0%	89.5%	94.5%	93.0%	89.1%
	Average width of HPD set	0.28	0.22	0.48	0.19	0.59

Note: SMRs in the last column is averaged over all the 100 sites.

Table 2: Comparison of the CK, YL, HB and AB approaches

Method	Parameter (True Value)	$\beta_0(0.1)$	$\beta_1(0.3)$	$\gamma(0.95)$	$\sigma^2(0.2)$
CK	Bias$\times 10^3$	-19.20	22.00	-211.99	-7.54
	MSE $\times 10^3$	8.84	4.38	90.28	2.27
	RMAD$\times 10^2$	77.65	17.56	22.84	19.61
	Average of Medians	0.05	0.32	0.79	0.19
YL	Bias$\times 10^3$	-51.74	29.07	-114.08	0.24
	MSE $\times 10^3$	12.37	5.58	61.55	5.39
	RMAD$\times 10^2$	98.18	19.62	16.33	26.05
	Average of Medians	0.01	0.33	0.93	0.19
	Coverage	27.0%	89.0%	—	—
HB	Bias$\times 10^3$	-46.20	27.25	-200.42	4.57
	MSE$\times 10^3$	10.99	4.83	63.51	2.54
	RMAD$\times 10^2$	91.14	18.48	21.15	19.71
	Average of MAPs	0.06	0.32	0.86	0.18
	Average of Medians	0.06	0.33	0.79	0.20
	Coverage	84.0%	89.5%	94.5%	93.0%
AB1	Bias$\times 10^3$	-47.03	27.42	-282.62	25.70
	MSE$\times 10^3$	11.17	4.84	126.26	3.08
	RMAD$\times 10^2$	92.05	18.49	29.84	21.32
	Average of MAPs	0.05	0.33	0.75	0.21
	Average of Medians	0.06	0.33	0.70	0.23
	Coverage	34.0%	82.5%	35.5%	80.0%
AB2	Bias$\times 10^3$	-46.83	27.16	-304.79	27.95
	MSE$\times 10^3$	11.21	4.83	140.45	3.19
	RMAD$\times 10^2$	92.19	18.46	32.13	21.88
	Average of MAPs	0.06	0.33	0.70	0.23
	Average of Medians	0.05	0.33	0.76	0.21
	Coverage	50.25%	90.45%	58.29%	94.47%

AB1: $\eta_z = 1$, $\eta_\pi = 5$, AB2: $\eta_z = 0.5$, $\eta_\pi = 20$

RMAD = relative mean absolute deviation

— values have not been reported by Yasui and Lele (1997)

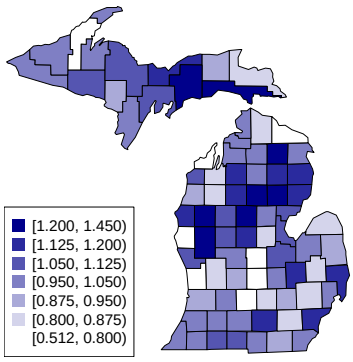
Table 3: Results for the Scottish Lip Cancer data

Method/Parameter	β_0	β_1(AFF/10)	γ	σ^2
CK	-0.25	0.40	0.99	0.47
YL	-0.14	0.08	1	1.12
(se)	(0.25)	(0.23)	—	—
HB	-0.31	0.38	0.96	0.64
(se)	(0.39)	(0.13)	(0.04)	(0.23)
MAP	-0.33	0.39	0.99	0.54
Median	-0.31	0.38	0.97	0.60
(HPD set)	(-0.92, 0.33)	(0.17, 0.59)	(0.92, 1.00)	(0.29, 0.98)
AB1	-0.29	0.37	0.96	0.61
(se)	(0.21)	(0.11)	(0.02)	(0.16)
MAP	-0.31	0.37	0.98	0.53
Median	-0.28	0.37	0.97	0.60
(HPD set)	(-0.63, 0.03)	(0.19, 0.54)	(0.93, 0.99)	(0.36, 0.85)
AB2	-0.31	0.37	0.95	0.67
(se)	(0.29)	(0.13)	(0.04)	(0.23)
MAP	-0.34	0.39	0.98	0.52
Median	-0.27	0.39	0.97	0.60
(HPD set)	(-0.77, 0.14)	(0.16, 0.58)	(0.90, 1.00)	(0.31, 1.00)

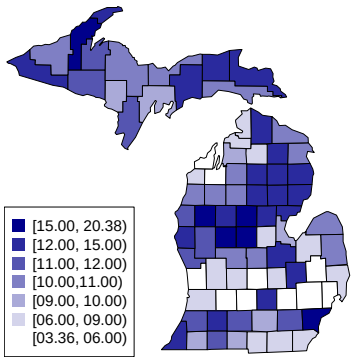
$AB1:$ $\eta_z = 1$, $\eta_\pi = 5$, AB2:$\eta_z = 0.5$, $\eta_\pi = 20$

— values have not been reported in Yasui and Lele (1997)

Figure 2: Observed SMR of Lung Cancer and Poverty in Michigan



(a) Obs SMR



(b) Poverty(%)

Table 4: The Analysis of Michigan Lung Cancer data

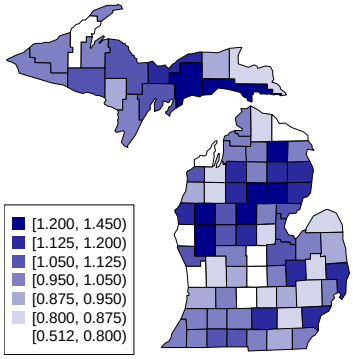
Method/Parameter	β_0	β_1 (poverty)	γ	σ^2
CK	-0.25	2.31	-0.60	0.06
YL	-0.27	2.51	-0.60	0.06
(sd)	(0.06)	(0.59)	*	*
HB	-0.23	2.19	-0.31	0.07
(sd)	(0.07)	(0.63)	(0.42)	(0.02)
MAP	-0.26	2.27	-0.55	0.06
Median	-0.24	2.21	-0.36	0.07
(HPD set)	(-0.35, -0.13)	(1.20, 3.22)	(-0.97, 0.30)	(0.04,0.09)
AB1	-0.22	2.07	-0.25	0.07
(sd)	(0.04)	(0.32)	(0.30)	(0.01)
MAP	-0.23	2.09	-0.30	0.06
Median	-0.22	2.09	-0.25	0.06
(HPD set)	(-0.28, -0.17)	(1.54,2.59)	(-0.74,0.22)	(0.05,0.09)
AB2	-0.23	2.12	-0.22	0.07
(sd)	(0.05)	(0.42)	(0.36)	(0.02)
MAP	-0.23	2.11	-0.39	0.06
Median	-0.23	2.17	-0.24	0.07
(HPD set)	(-0.31, -0.16)	(1.43,2.80)	(-0.80,0.34)	(0.04,0.09)
AB3	-0.24	2.19	-0.26	0.07
(sd)	(0.05)	(0.42)	(0.39)	(0.02)
MAP	-0.24	2.15	-0.43	0.06
Median	-0.23	2.21	-0.29	0.07
(HPD set)	(-0.31, -0.17)	(1.51,2.87)	(-0.90,0.30)	(0.04,0.09)

$AB1 : \eta_z = 1, \eta_\pi = 5$

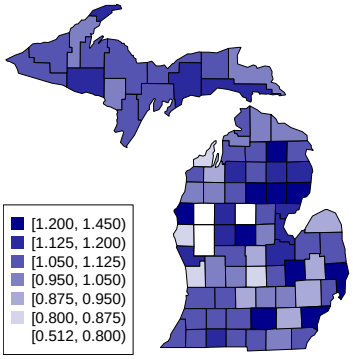
$AB2 : \eta_z = 0.5, \eta_\pi = 20$

$AB3 : \eta_z = (0.5, 0.5, 0.25, 0.125), \eta_\pi = 20$

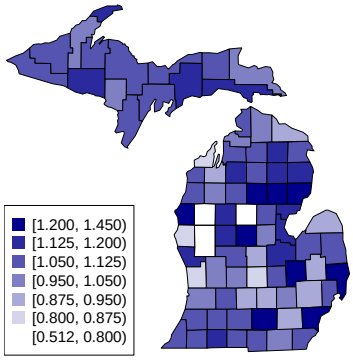
Figure 3: Observed SMR and estimated SMR of Lung Cancer in Michigan



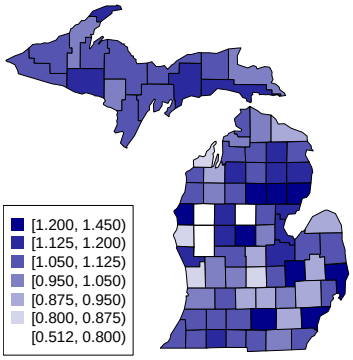
(a) Obs SMR



(b) EB SMR



(c) HB SMR



(d) AB3 SMR