Component of Statistics Canada
Catalogue no. 12-001-X Business Survey Methods Division

Article

Hierarchical Bayes small area estimation under a spatial model with application to health survey data

by Yong You and Qian M. Zhou

June 2011

Statistics Canada Statistique Canada

Canada

Survey Methodology, June 2011
Vol. 37, No. 1, pp. 25-37
Statistics Canada, Catalogue No. 12-001-X

Hierarchical Bayes small area estimation under a spatial model with application to health survey data

Yong You and Qian M. Zhou [1]

Abstract

In this paper we study small area estimation using area level models. We first consider the Fay-Herriot model (Fay and Herriot 1979) for the case of smoothed known sampling variances and the You-Chapman model (You and Chapman 2006) for the case of sampling variance modeling. Then we consider hierarchical Bayes (HB) spatial models that extend the Fay-Herriot and You-Chapman models by capturing both the geographically unstructured heterogeneity and spatial correlation effects among areas for local smoothing. The proposed models are implemented using the Gibbs sampling method for fully Bayesian inference. We apply the proposed models to the analysis of health survey data and make comparisons among the HB model-based estimates and direct design-based estimates. Our results have shown that the HB model-based estimates perform much better than the direct estimates. In addition, the proposed area level spatial models achieve smaller CVs than the Fay-Herriot and You-Chapman models, particularly for the areas with three or more neighbouring areas. Bayesian model comparison and model fit analysis are also presented.

Key Words: Area level model; Bayesian model comparison; Disease rate; Gibbs sampling; Hierarchical spatial model; Posterior predictive model checking; Sampling variance.

1. Introduction

Model-based small area estimation methods have been widely used in practice due to the increasing demand for precise estimates for local regions and various small areas. In general sample surveys are designed to provide reliable estimates for large regions or aggregates of small areas such as the whole nation and provinces. Direct survey estimates, based only on the area specific sample data, usually provide reliable estimates of the parameter of interest for those large areas. For small areas, particularly some small geographical areas or specific small domains, direct estimates are likely to yield large standard errors because of the small sample sizes in those small areas. Therefore in making inference for small areas, it is necessary to borrow strength from related areas to form indirect estimates that increase the effective sample size and thus increase the precision of estimates. It is now generally accepted that the indirect estimates should be based on explicit models that provide links to related areas through the use of supplementary data such as census counts or administrative records; see, for example, Rao (2003) and Jiang and Lahiri (2006) for more discussion on model-based small area methods. The model-based estimates are obtained to improve the direct design-based estimates in terms of precision and reliability, *i.e.*, smaller coefficients of variation (CVs). There are two broad classifications for small area models: area level models and unit level models. Area level models are based on area direct survey estimates and unit level models are based on individual observations in small areas. In this paper we focus on area level models that borrow strength across regions to improve the direct survey estimates.

Among the area level models, the Fay-Herriot model (Fay and Herriot 1979) is a basic and widely used area level model in practice to obtain reliable model-based estimates for small areas. The Fay-Herriot model basically has two components, namely, a sampling model for the direct estimates and a linking model for the parameters of interest. The sampling model involves the direct survey estimate and the corresponding sampling variance. The Fay-Herriot model assumes that the sampling variance is known in the model. Typically a smoothed estimator of the sampling variance is obtained and then treated as known in the model. Wang and Fuller (2003) and You and Chapman (2006) considered the situation where the sampling variances are unknown and modeled separately by direct estimators. In this paper we will consider both the smoothing and modeling methods for the sampling variances in the sampling model.

The linking model relates the parameter of interest to a regression model with area-specific random effects. In the Fay-Herriot model, the area random effects are usually assumed to be independent and identically distributed (*iid*) normal random variables to capture geographically unstructured variations among areas. However, in some small area applications, particularly in public health estimation problems, geographical variation of a disease is a subject of interest, and estimation of overall spatial pattern of risk and borrowing strength across regions to reduce variances of final estimates are both important. Thus, it may be more reasonable to construct spatial models on the area-specific

1. Yong You, Statistical Research and Innovation Division, Statistics Canada. E-mail: yongyou@statcan.gc.ca; Qian M. Zhou, Department of Biostatistics, Harvard University.

random effects to capture the spatial dependence among them. The spatial models are generally used in health related small area estimation, and various spatial models have been proposed for small area estimation (*e.g.*, Cressie 1990; Ghosh, Natarajan, Stroud and Carling 1998; Maiti 1998; Ghosh, Natarajan, Walter and Kim 1999; He and Sun 2000; Moura and Migon 2002; Singh, Shukla and Kundu 2005; Souza, Moura and Migon 2009). Best, Richardson and Thomson (2005) provided a comprehensive review on spatial models for disease mapping. Rao (2003) also discussed several spatial small area models.

The objective of this paper is to consider spatial correlation small area models and illustrate the usefulness of these models through an application to health survey data. The paper is organized as follows. In section 2, we first study area level models including the Fay-Herriot model and spatial correlation linking models. Then in section 3 we propose hierarchical Bayes (HB) small area models with spatial correlation and obtain HB inference for small area parameters through the Gibbs sampling method. In section 4, we apply the proposed models to the analysis of small area data from the Canadian Community Health Survey. We compare the performance of the model-based estimates with the direct design-based estimates, and moreover, we compare the proposed models with the Fay-Herriot model and the You-Chapman model (You and Chapman 2006) to investigate the effects of incorporating spatial structure on the area-specific random effects. Bayesian model comparison and model fit analysis are also provided. Finally in section 5, we offer some concluding remarks.

2. Small area models and inference

2.1 Fay-Herriot model

Let θ_i denote the parameter of interest for the i^{th} area, where $i = 1, ..., m$, and m is the total number of areas. The Fay-Herriot model assumes that the θ_i's are related to area specific auxiliary data $\mathbf{x}_i = (x_{i1}, ..., x_{ip})'$ through a linear regression model as follows:

$$\theta_i = \mathbf{x}_i'\beta + v_i,\ i = 1, ..., m \qquad (1)$$

where $\beta = (\beta_1, ..., \beta_p)'$ is the $p \times 1$ vector of regression coefficients, and the v_i's are area-specific random effects assumed to be *iid* with $\mathrm{E}(v_i) = 0$ and $\mathrm{Var}(v_i) = \sigma_v^2$. The assumption of normality may also be included. This model is referred to as a linking model for θ_i. The Fay-Herriot model also assumes that a direct survey estimator y_i, which is usually design-unbiased for the parameter of interest θ_i, is available whenever the area sample size $n_i > 1$. It is customary to assume that

$$y_i = \theta_i + e_i,\ i = 1, ..., m \qquad (2)$$

where e_i's are the sampling errors associated with the direct estimator y_i. We also assume that the e_i's are independent normal random variables with mean $E(e_i \,|\, \theta_i) = 0$ and sampling variance $\mathrm{Var}(e_i \,|\, \theta_i) = \sigma_i^2$. The model (2) is referred to as a sampling model for the direct survey estimator y_i. Combining these two components (1) and (2) leads to a linear mixed effects model (the Fay-Herriot model) as

$$y_i = \mathbf{x}_i'\beta + v_i + e_i,\ i = 1, ..., m. \qquad (3)$$

In the basic Fay-Herriot model (3), the sampling variances σ_i^2 are usually assumed as known, which is a very strong assumption. Generally, we can use direct sampling variance estimates from the survey data, however, these direct estimates are unstable if sample sizes are small. Therefore, in practice, a smoothed estimator of σ_i^2 is used in the model and treated as known. A generalized variance function is usually applied in practice to obtain a smoothed estimator for the sampling variance, *e.g.*, Dick (1995). In recent years, a method of smoothing design effects has been developed and used in practice to obtain smoothed variance estimators (*e.g.*, Singh, Folsom and Vaish 2005; You 2008a; Liu, Lahiri and Kalton 2008). In particular, You (2008a) applied an equal design effects modeling approach to obtain smooth estimates of sampling variances. The design effect for the i^{th} area may be approximately written as

$$\mathrm{deff}_i = \frac{s_i^2}{s_{ri}^2},\ \text{for } i = 1, ..., m,$$

where s_i^2 is the unbiased direct estimate of sampling variance based on the complex sampling design, and s_{ri}^2 is the estimate of sampling variance based on the assumption of simple random sampling design. For each area, based on the assumption of a common design effect, a smoothed factor deff can be obtained by $\mathrm{deff} = \sum_{i=1}^m \mathrm{deff}_i/m$. Then a smoothed sampling variance estimate $\tilde{\sigma}_i^2$ can be obtained as $\tilde{\sigma}_i^2 = s_{ri}^2 \cdot \mathrm{deff}$.

Instead of plugging in the smoothed estimates of sampling variances in the model, alternatively we can model the sampling variance directly. In the papers by Wang and Fuller (2003) and You and Chapman (2006), they assume the sampling variance σ_i^2 unknown and estimate σ_i^2 by an unbiased direct estimator s_i^2, which is independent of the direct survey estimator y_i. They also assume that $d_i s_i^2 \sim \sigma_i^2 \chi_{d_i}^2$, where $d_i = n_i - 1$, and n_i is the sample size for the i^{th} area. You and Chapman (2006) considered the full HB approach with the Gibbs sampling method which automatically takes into account the extra uncertainty associated with the estimation of σ_i^2. In this paper, we consider both the smoothing and modeling approaches for the sampling variances.

2.2 Spatial models

To incorporate spatially correlated random effects in the linking model, a simple and obvious way is to add a spatial random effect u_i in the independent linking model (1) as follows:

$$\theta_i = \mathbf{x}_i'\beta + v_i + u_i, \qquad (4)$$

where u_i's follow the well known intrinsic conditional autoregressive model given as

$$u_i \mid u_{-i} \sim \mathrm{N}\left(\frac{\sum_{j \neq i} w_{ij} u_j}{\sum_{j \neq i} w_{ij}}, \frac{\sigma_u^2}{\sum_{j \neq i} w_{ij}} \right), \qquad (5)$$

where u_{-i} denotes the values of spatial random effects u_j's in all other areas with $j \neq i$, weights w_{ij} are fixed constants, and σ_u^2 is a unknown variance component. In practice, a common choice of w_{ij} is to let $w_{ij} = 0$ unless areas i and j are neighboring areas (*i.e.*, share a common boundary), in which case $w_{ij} = 1$. The model (4) is proposed by Besag, York and Mollie (1991) to separate spatial effects from overall heterogeneity in the areas. In model (4), independent random effects v_i capture geographically unstructured heterogeneity among areas, and spatial random effects u_i capture spatial dependence between areas. In this way, the degree of overall spatial dependence can be expressed based on the proportion of the total variation in $v_i + u_i$ captured by each component.

In practice, it is often unclear how to choose between an unstructured model (*e.g.*, the basic linking model) given by (1) and a purely spatially structured model (*e.g.*, intrinsic autoregressive model) given by (5). For model (4), posterior inference about the spatial dependence is based on the proportion of the total variation in the sum of $v_i + u_i$ captured by each component. However, although the univariate conditional distributions of the spatial component (5) are well defined, the corresponding joint distribution is improper (with undefined mean and infinite variance). Moreover, the model (4) has a potential identifiability problem where only the sum of the random effects $v_i + u_i$ is well identified by the data; see, for example, Best *et al.* (2005), for a more detailed discussion.

Alternatively, we can consider another spatial parameterization studied by Leroux, Lei, and Breslow (1999) and MacNab (2003), which avoids the identifiability problem encountered with the model (4). Let $\theta_i = \mathbf{x}_i'\beta + b_i$, and $\mathbf{b} = (b_1, ..., b_m)'$. Following Leroux *et al.* (1999) and MacNab (2003), we place the following conditional autoregressive (CAR) model on the area specific spatial effects $\mathbf{b} = (b_1, ..., b_m)'$:

$$\mathbf{b} \sim \mathrm{MVN}(\mathbf{0}, \Sigma(\sigma_b^2, \lambda)) \qquad (6)$$

$$\Sigma(\sigma_b^2, \lambda) = \sigma_b^2 \mathbf{D}^{-1}, \ \mathbf{D} = \lambda \mathbf{R} + (1 - \lambda)\mathbf{I} \qquad (7)$$

where σ_b^2 is a spatial dispersion parameter and λ is a spatial autocorrelation parameter, $0 \leq \lambda \leq 1$; $\mathbf{I}$ is an identity matrix of dimension m; $\mathbf{R}$, commonly known as the neighbourhood matrix, has i^{th} diagonal element equal to the number of neighbors of the area i, and the off-diagonal elements in each row equal to -1 if the corresponding areas are neighbors and 0 otherwise. The CAR model (6) - (7) corresponds to the following conditional distribution of b_i:

$$b_i \mid b_{-i} \sim \mathrm{N}\left(\frac{\lambda}{1 - \lambda + \lambda w_{i+}} \sum_{j \neq i} w_{ij} v_j, \frac{\sigma_b^2}{1 - \lambda + \lambda w_{i+}} \right),$$

where $w_{i+} = \sum_{j \neq i} w_{ij}$. The CAR model (6) - (7) becomes the intrinsic autoregressive model (5) if $\lambda = 1$. On the other hand, if $\lambda = 0$, the CAR model (6) - (7) reduces to the independent linking model (1) which assumes independence on the area-specific random effects v_i. It is necessary to point out that the conditional mean and variances of $b_i \mid b_{-i}$ are weighted sums of the corresponding overall smoothing moments from the basic linking model (1) and local smoothing moments from the intrinsic autoregressive model:

$$\mathrm{E}(b_i \mid b_{-i}) = \frac{1 - \lambda}{1 - \lambda + \lambda w_{i+}} \times 0 + \frac{\lambda w_{i+}}{1 - \lambda + \lambda w_{i+}} \left(\sum_{j \neq} w_{ij} b_j / w_{i+} \right)$$

$$\mathrm{Var}(b_i \mid b_{-i}) = \frac{1 - \lambda}{1 - \lambda + \lambda w_{i+}} \times \sigma_b^2 + \frac{\lambda w_{i+}}{1 - \lambda + \lambda w_{i+}} (\sigma_b^2 / w_{i+}).$$

Thus model (6)-(7) is a balance between the independent linking model (1) and the intrinsic CAR model (5). The spatial correlation parameter λ measures the extent of the spatial effects for local smoothing of the neighbouring areas. The modeling structure (6) captures both the unstructured heterogeneity among areas and the spatial correlation effects of the neighbouring area.

2.3 Hierarchical Bayes models and inference

In order to estimate θ_i, the parameter of interest, we apply a hierarchical Bayes (HB) approach using the Gibbs sampling method. Compared to other approaches such as EBLUP and empirical Bayes (EB), HB approach is straightforward and the inference for θ_i are exact unlike the EB or EBLUP. Moreover, the HB approach can deal with complex small area models using the Monte Carlo Markov Chain

(MCMC) method, which overcomes the computational difficulties of multi-dimensional integrations of posterior quantities to a large extent.

Let $\mathbf{y} = (y_1, ..., y_m)'$, $\boldsymbol{\theta} = (\theta_1, ..., \theta_m)'$, and $\mathbf{X} = (\mathbf{x}_1, ..., \mathbf{x}_m)'$. We first construct two HB models without and with spatial structure under the assumption that the sampling variance σ_i^2 are assumed known and replaced by the smoothed estimate $\tilde{\sigma}_i^2$.

Model 1: Fay-Herriot model, denoted as FHM (Fay and Herriot 1979; Rao 2003).

- $y_i \mid \theta_i \sim \mathrm{N}(\theta_i, \sigma_i^2 = \tilde{\sigma}_i^2)$, for $i = 1, ..., m$;
- $\theta_i \mid \beta, \sigma_v^2 \sim \mathrm{N}(\mathbf{x}_i'\beta, \sigma_v^2)$, for $i = 1, ..., m$;
- Priors for the parameters (β, σ_v^2): $\pi(\beta) \propto 1$; $\pi(\sigma_v^2) \sim \mathrm{IG}(a_0, b_0)$, where a_0, b_0 are chosen to be very small known constants to reflect vague knowledge on σ_v^2. N stands for the normal distribution and IG for the inverse gamma distribution.

Model 2: Proposed area level CAR model, as an extension of the Fay-Herriot model, denoted as CAR-FHM.

- $\mathbf{y} \mid \boldsymbol{\theta} \sim \mathrm{MVN}(\boldsymbol{\theta}, \mathbf{E})$, where $\mathbf{E}$ is a diagonal matrix with the i^{th} diagonal element $\sigma_i^2 = \tilde{\sigma}_i^2$;
- $\boldsymbol{\theta} \mid \beta, \sigma_v^2 \sim \mathrm{MVN}(\mathbf{X}\beta, \sigma_v^2\mathbf{D}^{-1})$, where $\mathbf{D} = \lambda\mathbf{R} + (1 - \lambda)\mathbf{I}$, with $\mathbf{I}$, an identity matrix of dimension m, and $\mathbf{R}$, the neighbourhood matrix;
- Priors for the parameters $(\beta, \lambda, \sigma_v^2)$: $\pi(\beta) \propto 1$; $\pi(\lambda) \sim \mathrm{Uniform}(0, 1)$, where $0 \le \lambda \le 1$; $\pi(\sigma_v^2) \sim \mathrm{IG}(a_0, b_0)$, where a_0, b_0 are chosen to be very small known constants. MVN stands for the multivariate normal distribution.

Note that the proposed model CAR-FHM reduces to FHM when the spatial autocorrelation parameter $\lambda = 0$.

We also consider two HB models with the sampling variance σ_i^2 unknown and modeled by the direct unbiased estimator s_i^2.

Model 3: You-Chapman Model, denoted as YCM (You and Chapman 2006).

- $y_i \mid \theta_i, \sigma_i^2 \sim \mathrm{N}(\theta_i, \sigma_i^2)$, for $i = 1, ..., m$;
- $d_i s_i^2 \mid \sigma_i^2 \overset{\text{ind}}{\sim} \sigma_i^2 \chi_{d_i}^2$ where $d_i = n_i - 1$, for $i = 1, ..., m$;
- $\theta_i \mid \beta, \sigma_v^2 \sim \mathrm{N}(\mathbf{x}_i'\beta, \sigma_v^2)$, for $i = 1, ..., m$;
- Priors for unknown parameters $(\beta, \sigma_v^2, \sigma_i^2, i = 1, ..., m)$: $\pi(\beta) \propto 1$; $\pi(\sigma_v^2) \sim \mathrm{IG}(a_0, b_0)$, $\pi(\sigma_i^2) \sim \mathrm{IG}(a_i, b_i)$ for $i = 1, ..., m$, where a_i, b_i $(0 \le i \le m)$ are chosen to be very small known constants to reflect vague knowledge on σ_i^2 and σ_v^2.

Model 4: Proposed area level CAR model with unknown sampling variances, as an extension of You-Chapman model, denoted as CAR-YCM.

- $\mathbf{y} \mid \boldsymbol{\theta}, \sigma_1^2, ..., \sigma_m^2 \sim \mathrm{MVN}(\boldsymbol{\theta}, \mathbf{E})$, where matrix $\mathbf{E}$ has diagonal elements σ_i^2;
- $d_i s_i^2 \mid \sigma_i^2 \overset{\text{ind}}{\sim} \sigma_i^2 \chi_{d_i}^2$, where $d_i = n_i - 1$, for $i = 1, ..., m$;
- $\boldsymbol{\theta} \mid \beta, \sigma_v^2 \sim \mathrm{MVN}(\mathbf{X}\beta, \sigma_v^2 \mathbf{D}^{-1})$, where $\mathbf{D} = \lambda\mathbf{R} + (1 - \lambda)\mathbf{I}$;
- Priors for the parameters $(\beta, \lambda, \sigma_v^2, \sigma_i^2, i = 1, ..., m)$: $\pi(\beta) \propto 1$; $\pi(\lambda) \sim \mathrm{Uniform}(0, 1)$, where $0 \le \lambda \le 1$; $\pi(\sigma_v^2) \sim \mathrm{IG}(a_0, b_0)$; $\pi(\sigma_i^2) \sim \mathrm{IG}(a_i, b_i)$ for $i = 1, ..., m$, where a_i, b_i $(0 \le i \le m)$ are chosen to be very small known constants.

Again, note that the proposed model CAR-YCM reduces to the You-Chapman model when $\lambda = 0$. For both models 3 and 4 there is an implicit assumption that the area-specific sample size $n_i \ge 2$. If flat priors are used for σ_i^2, we should have $n_i \ge 4$ to ensure proper posteriors (You and Chapman 2006).

We apply the Gibbs sampling method to estimate the posterior mean $\mathrm{E}(\theta_i \mid \mathbf{y})$ and the corresponding posterior variance $\mathrm{Var}(\theta_i \mid \mathbf{y})$. The required full conditional distributions of parameters under different models are given in Appendix A. For the Fay-Herriot model and the You-Chapman model, all the full conditional distributions have closed forms and drawing samples from these distributions is straightforward. For the proposed two area level spatial models CAR-FHM and CAR-YCM, the conditional distribution of the spatial correlation parameter λ does not have a closed form. We use the Metropolis-Hastings algorithm within the Gibbs sampler (Chip and Greenberg 1995) to update λ. Under the model CAR-FHM, the full conditional distribution of λ in the Gibbs sampler can be written as

$$[\lambda \mid \boldsymbol{\theta}, \beta, \sigma_v^2] \propto \mathrm{h}(\lambda)\,\mathrm{f}(\lambda)$$

where $\mathrm{f}(\lambda)$ is a density function of the uniform distribution, Uniform (0, 1), given as

$$\mathrm{f}(\lambda) \propto 1, \text{ where } 0 \le \lambda \le 1$$

and $\mathrm{h}(\lambda)$ is a function given by

$$\mathrm{h}(\lambda) \propto \left|\left[\lambda\mathbf{R} + (1 - \lambda)\mathbf{I}\right]^{-1}\right|^{-1/2} \times \exp\left\{-\frac{1}{2\sigma_v^2}(\boldsymbol{\theta} - \mathbf{X}\beta)'\left[\lambda\mathbf{R} + (1 - \lambda)\mathbf{I}\right](\boldsymbol{\theta} - \mathbf{X}\beta)\right\}.$$

We use $f(\lambda)$ as the "candidate" generating density function in the Metropolis-Hastings updating step. To update λ from the current values of $(\boldsymbol{\theta}^{(k)}, \beta^{(k)}, \sigma_v^{2(k)})$, we proceed as follows:

1. Draw λ^* from a uniform distribution;
2. Compute the acceptance probability $\alpha(\lambda^*, \lambda^{(k)}) = \min\{h(\lambda^*)/h(\lambda^{(k)}), 1\}$;
3. Generate u from a uniform distribution, if $u < \alpha(\lambda^*, \lambda^{(k)})$, then the candidate value λ^* is accepted, *i.e.*, $\lambda^{(k+1)} = \lambda^*$; otherwise λ^* is rejected, and set $\lambda^{(k+1)} = \lambda^{(k)}$.

For the model CAR-YCM, a similar procedure can be applied when drawing samples from the conditional distribution of λ.

3. Data analysis

3.1 Data description and implementation

The Canadian Community Health Survey (CCHS) is a federal survey conducted by Statistics Canada. The primary objective of CCHS is to provide timely and reliable estimates of health determinants, health status and health system utilization across Canada. It is a cross-sectional survey which operates on a two-year collection cycle. The first year of the survey cycle "x.1" targets individuals aged 12 or older who are living in private dwellings, and it is a general population health survey with a large sample (130,000 persons) designed to provide reliable estimates at the health region, provincial and national levels. The second year of the survey cycle "x.2" has a smaller sample (30,000 persons) allocated based on provincial sample buy-ins and is designed to provide provincial and national level results on specific focused health topics. Although national and provincial estimates are very important, there is an increasing demand for health data at lower levels of geography voiced by a number of provinces including British Columbia (BC), Prince Edward Island (PEI), Quebec and others. Cycle "x.1" of the CCHS collected data corresponds to 136 health regions in the 10 provinces and three territories. It primarily used two sampling frames. The first one, used as the primary frame, was based on the area frame designed for the Canadian Labour Force Survey, and within the area frame, a multistage stratified cluster design was used to sample dwellings. The second frame consists of a list of telephone numbers. Random digit dialing methodology is used in some of the health regions for cost reasons. More details of the design are provided in Béland (2002). In this paper, we use a small data set from Cycle 1.1 as an example to demonstrate the analysis. We are interested in estimating the disease rate for local health regions within provinces. In particular, we apply the four models discussed in section 2 to estimate the asthma rate for 20 health regions in the province of BC using the data from Cycle 1.1. Figure 1 shows the map of the 20 health regions in the province of British Columbia. We use this map to define the neighbourhood correlation matrix used in the spatial models. Appendix B gives the list of health regions and related spatial structures.

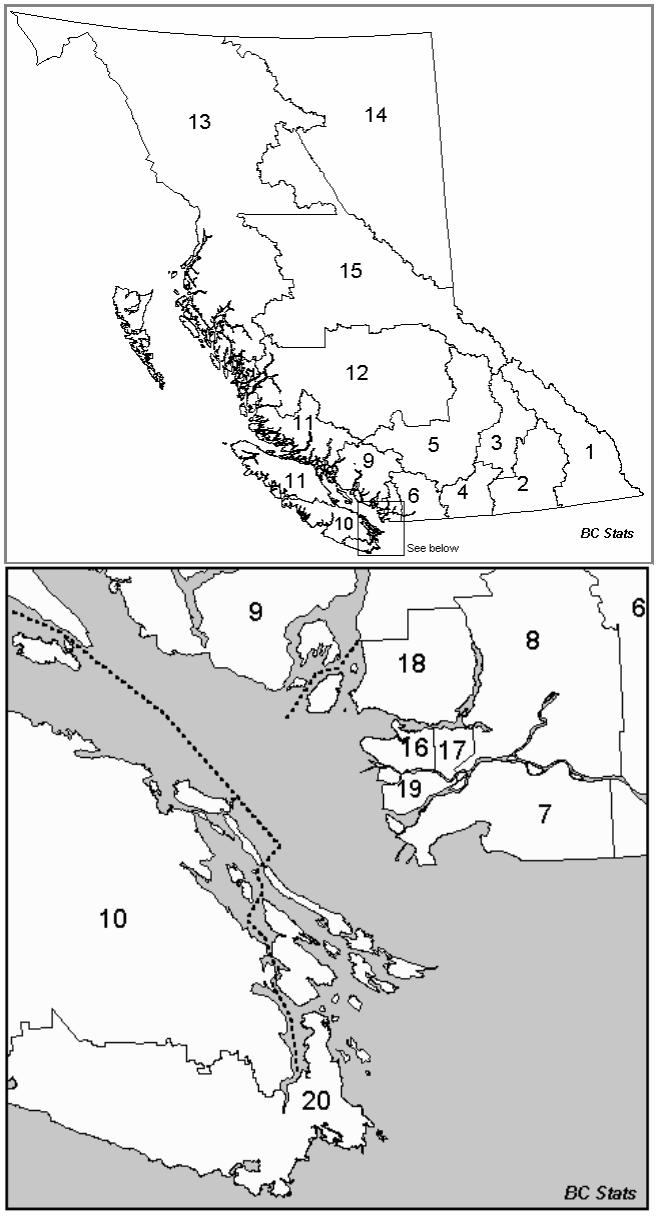

Figure 1 Map of 20 health regions in the province of British Columbia

Let θ_i denote the true asthma rate for the i^{th} health region in BC, $i = 1, \ldots, 20$. From the survey data of Cycle 1.1, we obtained the direct survey estimate y_i of θ_i as the ratio of number of people having asthma (direct survey

estimate) divided by the corresponding population size (known constant). We have also included six area level auxiliary variables used in the model, and these six variables are total population size, number of persons who have asthma as one of the symptoms of the chronic disease, number of persons who have asthma as the main symptom of the chronic disease, number of persons who have diabetes as one of the symptoms of the chronic disease, number of persons who have diabetes as the main symptom of the chronic disease, and number of visits to hospitals. Note that in the literature related to disease mapping (*e.g.*, Mollié 1996; Maiti 1998; MacNab 2003), a Poisson or Binomial distribution is usually assumed in the sampling model for the direct estimate y_i. However, in small area estimation, the direct estimate y_i is obtained based on the complex sampling design used in the survey. Thus, it is a customary approach to assume a normal sampling model on the direct estimates y_i; see, for example, Datta, Lahiri, Maiti and Lu (1999), Rao (2003), Mohadjer, Rao, Liu, Krenzke and Van de Kerckhove (2007), and You (2008a) . Note that we have only considered one kind of disease rate data from one province in our study and used this example as illustration of the proposed model and evaluate the effects of spatial modeling in small area models.

To implement the Gibbs sampling, we use $L = 5$ parallel runs each with a "burn-in" length of $B = 2{,}000$ and Gibbs sampling size of $G = 5{,}000$. For the proposed models CAR-FHM and CAR-YCM, in order to reduce the autocorrelation which results from the accept-rejection algorithm in the run, we take every 5^{th} iteration after the "burn-in" period. Therefore, for models FHM and YCM, we have $n = 5{,}000$ samples for each run, and for models CAR-FHM and CAR-YCM, we have $n = 1{,}000$ samples for each run. Convergence of the Gibbs sampling is monitored for the small area parameters θ_i and other unknown parameters in the model using the potential scale reduction factor (Gelman and Rubin 1992; Gelman, Carlin, Stern and Rubin 2004, page 296-297). We have computed the reduction factors for all the monitored parameters in the model in the Gibbs sampling. These factor values are all very close to 1 (less than 1.05), which suggests that the desired convergence for these parameters is achieved by the Gibbs sampler.

We have used vague priors for the hyperparameters in the model as a common practice in HB small area estimation. In particular, the flat prior for regression parameter $\pi(\beta) \propto 1$ and proper inverse gamma priors for variance components are commonly used (*e.g.*, Arora and Lahiri 1997; Ghosh *et al.* 1998; Datta *et al.* 1999; You and Rao 2000; Rao 2003, page 237; Souza *et al.* 2009). Following MacNab (2003), we have used the uniform prior $\pi(\lambda) \sim \text{Uniform}(0, 1)$ for the autocorrelation parameter. The uniform priors are also commonly used for the autocorrelation parameters in spatial models (*e.g.*, Maiti 1998; He and Sun 2000; Rao 2003, page 266). We also tried several different values for the inverse gamma priors. The HB estimates are quite stable and not sensitive to the choice of vague proper priors. More detailed discussion on sensitivity analysis can be found, for example, in You and Chapman (2006) for similar models.

3.2 Comparison of results

At first, we present the HB estimates of the asthma rate under models FHM and CAR-FHM in which the sampling variances σ_i^2 are assumed to be known. We used the smoothed estimate $\tilde{\sigma}_i^2$ obtained by the smoothing technique in You (2008a) as described in Section 2. Figure 2 displays the direct estimates and the HB model-based estimates under FHM and CAR-FHM for the 20 health regions in BC. The health regions appear in the x-coordinate ranked by the order of sample size with the smallest (Peace Liard) on the left and the largest (South Fraser Valley) on the right. Model 1 (FHM) and Model 2 (CAR-FHM) give similar point estimates, and both the model-based estimates lead to moderate smooth estimates compared to the direct estimates. Moreover, the direct estimates and two HB estimates of the disease rate are very close for some health regions with large sample sizes, but for some areas with smaller sample sizes, they differ to some extent. Similar results are obtained under Model 3 (YCM) and Model 4 (CAR-YCM).

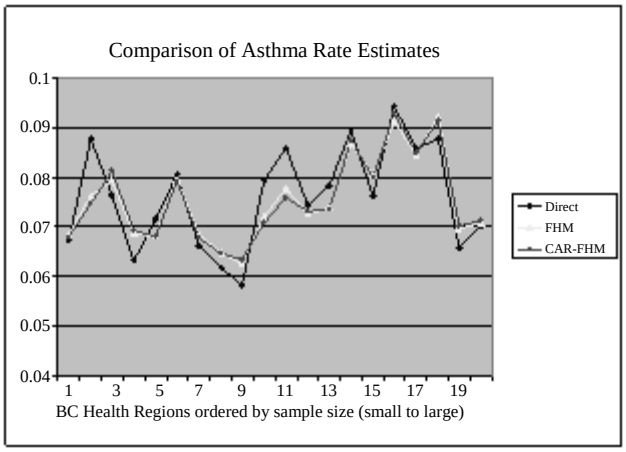

Figure 2 Direct and HB model-based estimates under models FHM and CAR-FHM

Figure 3 presents the CVs of the direct and two HB model-based estimates with the health regions ordered by the sample sizes from the smallest to the largest as in Figure 2. The CVs of HB estimates are obtained by dividing the squared root of the posterior variance by the posterior mean. As expected, the CVs of the direct estimates show a clear tendency of decrease as the sample size increases. However, the two model-based estimates give smoother CVs. Moreover, the two HB model-based estimates exhibit a great

improvement over the direct design-based estimates in terms of precision and reliability, that is, smaller CVs. Compared to the direct estimates, the average CV reduction of the HB estimates under FHM is about 22.7% ranging from 7.8% to 40.5%, and the average reduction of the CVs for the HB estimates under the proposed CAR-FHM is 27.8% ranging from 12.5% to 52.1%. Thus it is clear that the proposed spatial model CAR-FHM is superior to the Fay-Herriot model. We also obtained similar results for the models YCM and CAR-YCM when the sampling variance is modeled directly. The average CV reduction under YCM is 23.9%, whereas the average CV reduction is 29.0% under the proposed spatial model CAR-YCM. Details of the results including the point estimates and the corresponding CVs are presented in a table in Appendix C. In our example, the sample size at the health region level is relatively large. The model-based estimates have still shown great improvement over the direct survey estimates. Our results indicate that the presented small area models can be used to improve the direct survey estimates even when the sample size is relatively large. Note that Bayesian credible intervals for the small area parameters can be easily constructed using the MCMC output from the Gibbs sampler if required by practical users. This is an advantage of using the HB inference via MCMC sampling. However in this paper we only report the model-based point estimates and the corresponding CVs as our main purpose is to compare the model-based estimates with the direct estimates and to show the efficiency gain of the models. The gain in efficiency is clearly evident.

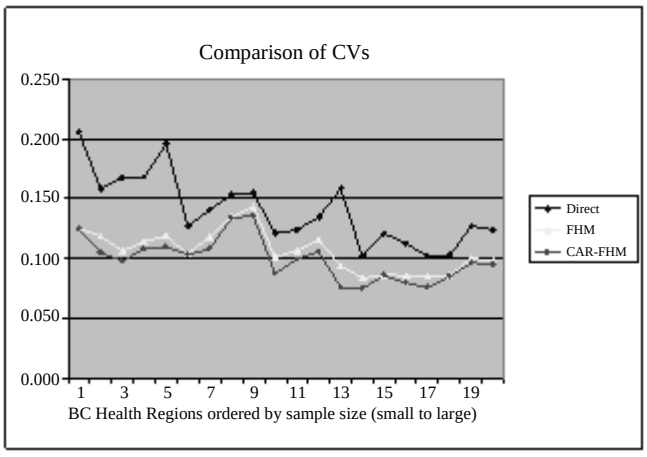

Figure 3 Direct and HB CVs under models FHM and CAR-FHM

In order to investigate the effects of incorporating the spatial structure in the model, we present the CVs of the direct and HB estimates by health regions sorted according to the number of neighbouring regions from the smallest (2 neighbours) to the largest (7 neighbours) in Figure 4. It shows that the HB estimates from the proposed model CAR-FHM has smaller CVs than the estimates from the Fay-Herriot model. In addition, the improvement of CAR-FHM over the Fay-Herriot model is much more obvious in the regions with more neighbours, and these two models give very close CVs in the regions with less adjacent areas. Very similar results are also obtained for CAR-YCM over YCM. Table 1 gives the average reduction of the CVs across the health regions with the same number of neighbours. The results in Table 1 present the CV reduction of the proposed spatial models for both cases of known and unknown sampling variances. For example, for known σ_i^2 (smoothed $\tilde{\sigma}_i^2$), for areas with only 2 neighbours, the average CV reduction of model CAR-FHM over the Fay-Herriot model is only around 0.9%, whereas for areas with 7 neighbours, the average CV reduction for CAR-FHM over FHM is as high as around 20%. For the case of unknown σ_i^2, similar results are obtained for CAR-YCM over YCM. The numerical results in Table 1 confirm the clear trend of increased CV reduction under the proposed spatial model over FHM or YCM as the number of neighbours increases. Thus, more neighbouring areas can provide more information in the spatial structure to improve the precision and reliability of the HB estimates.

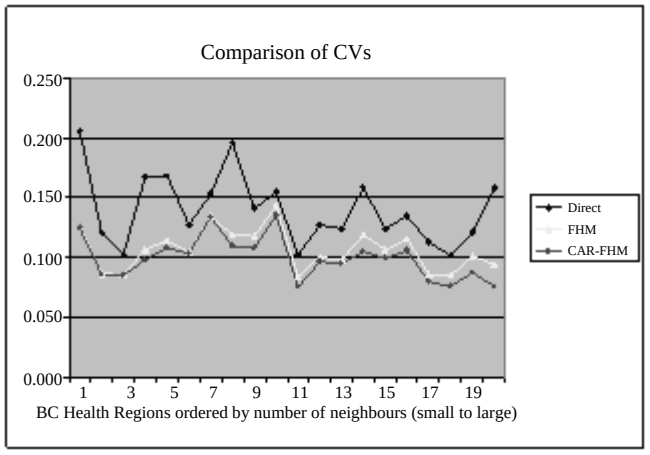

Figure 4 Direct and HB CVs under models FHM and CAR-FHM with the health regions sorted by the number of neighbours

Table 1
Comparison of average CV reduction

Number of neighbours	Average CV reduction	
	CAR-FHM over FHM	CAR-YCM over YCM
2	0.9%	1.8%
3	3.7%	3.5%
4	6.3%	6.0%
5	8.9%	8.7%
6	13.7%	11.0%
7	19.2%	20.7%

3.3 Bayesian model comparison

In this section, we compare the proposed models CAR-FHM with FHM and CAR-YCM with YCM, respectively. For hierarchical Bayes model comparison, the deviance information criterion (DIC) proposed by Spiegelhalter, Best, Carlin and van der Linde (2002) is commonly used in recent years to compare non-nested and mixed effects Bayesian models. The DIC is based on the deviance of the model $D(\theta)$, which is equal to minus twice the log-likelihood of the model, and the DIC is usually computed as DIC = $D(\hat{\theta}) + 2p_D$, where $D(\hat{\theta})$ is the deviance of the model evaluated at the posterior mean of the model parameters, which summarizes the goodness of fit of the model, and p_D is the effective number of parameters, which captures the complexity of the model. p_D is defined as $p_D = \bar{D}(\theta) - D(\hat{\theta})$, and $\bar{D}(\theta)$ is the posterior mean of the deviance of the model. Thus the DIC is defined as the summation of the goodness of fit of the model and the model complexity. Smaller values of DIC indicate a better model fit. Computation of DIC is relatively straightforward provided that the deviance $D(\theta)$ is available in closed form, and p_D may be calculated after the Gibbs sampling run by taking the sample mean of the simulated values of $D(\theta)$ minus the plug-in estimate of the deviance $D(\hat{\theta})$. For the four models presented in section 2, we computed the corresponding DIC values, as shown in Table 2. It is clear that the proposed spatial models CAR-FHM and CAR-YCM both have smaller DIC values than the non-spatial models FHM and YCM respectively, which indicates that the spatial models are better than the non-spatial models in our study. Both spatial models CAR-FHM and CAR-YCM perform well in this example. This result of model comparison is consistent with the estimation results presented in section 3.2.

Table 2
Comparison of DIC values for the four hierarchical models

Model	DIC value
FHM	27.1
CAR-FHM	24.6
YCM	26.8
CAR-YCM	24.5

3.4 Test of model fit

In order to check the overall model fit of the proposed models CAR-FHM and CAR-YCM, we use the method of posterior predictive distribution. Let y_{rep} denote the replicated observation under the model. The posterior predictive distribution of y_{rep} given the observed data y_{obs} is defined as $f(y_{\text{rep}} \mid y_{\text{obs}}) = \int f(y_{\text{rep}} \mid \theta)\, f(\theta \mid y_{\text{obs}})\, d\theta$. In this approach, a test statistic $T(y, \theta)$ that depends on the data y and possibly the parameter θ can be defined and the observed value $T(y_{\text{obs}}, \theta \mid y_{\text{obs}})$ compared to the posterior predictive distribution of $T(y_{\text{rep}}, \theta \mid y_{\text{obs}})$ with any signifycant difference indicates a model failure. Lack of fit of the data with respect to the posterior predictive distribution can be measured by the *p*-value of the test quantity (Meng 1994; Gelman, Meng and Stern 1996). The posterior predictive *p*-value is defined as $p = P(T(y_{\text{rep}}, \theta) \geq T(y_{\text{obs}}, \theta) \mid y_{\text{obs}})$. If the given model adequately fits the observed data, then $T(y_{\text{obs}}, \theta \mid y_{\text{obs}})$ should be near the central part of the histogram of the $T(y_{\text{rep}}, \theta \mid y_{\text{obs}})$ values if y_{rep} is generated repeatedly from the posterior predictive distribution. Consequently, the posterior predictive *p*-value is expected to be near 0.5 if the model adequately fits the data. Extreme *p*-values (near 0 or 1) suggest poor fit. The posterior predictive *p*-value model checking has been criticized for being conservative due to the double use of the observed data; see, for example, Bayarri and Berger (2000). They proposed alternative model checking *p*-value measures, named the partial posterior predictive *p*-value and the conditional predictive *p*-value. However, their methods are more difficult to implement and interpret (Rao 2003; Sinharay and Stern, 2003). As noted in Sinharay and Stern (2003), the posterior predictive *p*-value is especially useful if we think of the current model as a plausible ending point with modifications to be made only if substantial lack of fit is found.

To carry out the posterior predictive model checking, we need to specify a test quantity $T(y, \theta)$. You (2008b) studied several test quantities in posterior predictive model checking for small area models through a simulation study and proposed a test quantity given as

$$T(y, \theta) = \left| \max(y_i) - \text{mean}(\theta_i) \right| - \left| \min(y_i) - \text{mean}(\theta_i) \right|.$$

It is shown in You (2008b) that the proposed test quantity $T(y, \theta)$ is sensitive to the choice of distribution of random effects and different mean functions under the Fay-Herriot model. A similar test quantity is also suggested in Gelman *et al.* (2004) for posterior predictive model checking. In our study, under the proposed model CAR-FHM, the estimated *p*-value is 0.472, and under model CAR-YCM, the estimated *p*-value is 0.453. Thus there is no indication of lack of model fit and both proposed spatial models fit the data quite well.

To access model fit at the individual observation level, we also computed the individual predictive probability values p_i^* as $p_i^* = P(y_{i(\text{rep})} < y_{i(\text{obs})} \mid y_{\text{obs}})$; see, for example, Gelfand (1996) and Daniels and Gatsonis (1999). These individual predictive probabilities provide information on the degree of consistent overestimation or underestimation of the observed data. For model CAR-FHM, the p_i^* ranges from 0.325 to 0.768 with a mean of 0.517 and a median of 0.496; for model CAR-YCM, the p_i^* ranges from 0.316 to

0.772 with a mean of 0.511 and a median of 0.497. Both models give very similar results and the mean and median values are all around 0.5. There is no indication of any consistent overestimation or underestimation of the proposed models. The overall *p*-values and individual predictive probabilities have shown that the proposed spatial small area models fit the data quite well.

3.5 Bias diagnostics

To evaluate any possible bias of the model-based estimates under the proposed models with respect to the direct survey estimates, following Brown, Chambers, Heady and Heasman (2001), we consider a simple method of regression analysis for the direct estimates and the HB model-based estimates. You (2008a) also used the regression analysis method for model bias diagnostics. If the model-based estimates are close to the true values of the small area disease rate, then the direct survey estimates, which are assumed to be unbiased for the true disease rates, should behave like random variables whose expected values correspond to the values of the model-based estimates. That means the model-based estimates should be unbiased predictors of the direct estimates. In terms of regression analysis, we basically fit the regression model $Y = \alpha + \beta X$ to the data and estimate the coefficients, and see how close the regression line is to $Y = X$. Let Y be the direct survey estimates and X be the model-based estimates. Under the proposed CAR-FHM, we obtain a regression line $Y = -0.0021(0.011) + 1.0365(0.1445)X$; under the proposed CAR-YCM, we obtain a regression line $Y = -0.0028(0.0108) + 1.0458(0.1427)X$. Thus both the regression lines show very little disparity from $Y = X$. We therefore conclude that the model-based estimates are consistent with the direct estimates with no extra possible bias induced by the proposed models. The results also provide an indication of no evidence of any bias due to possible model misspecification.

4. Conclusions

In this paper we have discussed two area level models, namely, the well-known Fay-Herriot model in which the sampling variance is assumed to be known, and the You-Chapman model in which the sampling variance is unknown and modeled separately by its direct estimator. In both the Fay-Herriot model and You-Chapman model, the area random effects are assumed to be *iid* normal random variables to capture unexplained area heterogeneity effects. After comparing various forms of Gaussian CAR models proposed in the literature (*e.g.*, Best *et al.* 2005) for disease mapping to incorporate spatially correlated effects, we extended the independent area effects model to a spatial correlation model and combined it with the traditional small area models. The proposed new small area spatial correlation models CAR-FHM and CAR-YCM include the small area sampling models and a spatial correlation linking model which captures both the unstructured heterogeneity among areas and the spatial correlation effects of the neighbouring areas. We don't need to specify the spatial autocorrelation parameter in the model, and this parameter will be estimated from the data.

In the data analysis we compared the proposed spatial models with the non-spatial effects models by applying the models to estimate the rates of asthma for 20 health regions in the province of British Columbia. Our results have shown that the model-based estimates achieve a great improvement over the direct estimates in terms of moderately smoothed point estimates and much smaller CVs. Particularly, the proposed models are superior to the Fay-Herriot model or You-Chapman model whether the sampling variances are assumed to be known or unknown. Moreover, note that the CV reduction of the proposed spatial models over the Fay-Herriot model or You-Chapman model is greater for the areas with more neighbours. Results of the Bayesian model comparison and model fit analysis are also in favor of the proposed small area spatial models.

In future work, the proposed small area spatial models can be extended to unmatched sampling and linking models (You and Rao 2002) with the sampling variance known or unknown. We plan to evaluate the estimation effects of different spatial models as well as the effects of spatial structures. For data analysis, we will produce model-based health status estimates based on the proposed models for health regions across Canada and evaluate the possibility of extending the model-based approach to lower level estimates such as age-sex domains within heath regions. We also plan to consider the data cloning method (Lele, Dennis and Lutscher 2007; Lele, Nadeem and Schmuland 2010) for the spatial models. An advantage of data cloning method is that the results are independent of the choice of priors. But the computational burden could be considerably extensive.

Acknowledgements

We would like to thank one Associate Editor and one referee for their detailed comments and suggestions. Yong You's research work was supported by Statistics Canada Methodology Branch Research Block Fund. Qian M. Zhou's work was finished when she worked at Statistics Canada as a MITACS/NPCDS research internship student under the supervision of Yong You. Q.M. Zhou presented the proposed models and some results of the paper at the 2008 Statistics Society of Canada (SSC) annual meeting in Ottawa, and won the 2008 best student paper award of the SSC Survey Methods Section.

Appendix A

Full conditional distributions

A.1. Gibbs sampling full conditional distributions under Model 1: FHM.

- $[\theta_i \mid y_i, \beta, \sigma_v^2] \sim \mathrm{N}[\gamma_i y_i + (1-\gamma_i)\mathbf{x}_i'\beta, \tilde{\sigma}_i^2\gamma_i]$, where $\gamma_i = \sigma_v^2/(\sigma_v^2+\tilde{\sigma}_i^2)$, for $i = 1, ..., m$;
- $[\beta|\boldsymbol{\theta}, \sigma_v^2] \sim \mathrm{N}\left[\left(\sum_{i=1}^m \mathbf{x}_i\mathbf{x}'\right)^{-1}\left(\sum_{i=1}^m \mathbf{x}_i\theta_i\right), \sigma_v^2\left(\sum_{i=1}^m \mathbf{x}_i\mathbf{x}_i'\right)^{-1}\right]$;
- $[\sigma_v^2 \mid \boldsymbol{\theta}, \beta] \sim \mathrm{IG}\left[a_0 + \frac{1}{2}m, b_0 + \frac{1}{2}\sum_{i=1}^m (\theta_i - \mathbf{x}_i'\beta)^2\right]$.

A.2. Gibbs sampling full conditional distributions under Model 2: CAR-FHM.

- $[\boldsymbol{\theta} \mid \mathbf{y}, \beta, \lambda, \sigma_v^2] \sim \mathrm{MVN}(\boldsymbol{\Lambda}\mathbf{y} + (\mathbf{I}-\boldsymbol{\Lambda})\mathbf{X}\beta, \boldsymbol{\Lambda}\mathbf{E})$, where $\boldsymbol{\Lambda} = (\mathbf{E}^{-1}+\mathbf{D}/\sigma_v^2)^{-1}\mathbf{E}^{-1}$ with $\mathbf{E} = \mathrm{diag}\{\tilde{\sigma}_1^2, ..., \tilde{\sigma}_m^2\}$ and $\mathbf{D} = \lambda\mathbf{R} + (1-\lambda)\mathbf{I}$;
- $[\beta|\boldsymbol{\theta}, \lambda, \sigma_v^2] \sim \mathrm{MVN}[(\mathbf{X}'\mathbf{D}\mathbf{X})^{-1}\mathbf{X}'\mathbf{D}\boldsymbol{\theta}, \sigma_v^2(\mathbf{X}'\mathbf{D}\mathbf{X})^{-1}]$;
- $[\lambda|\boldsymbol{\theta}, \beta, \sigma_v^2] \propto \left|[\lambda\mathbf{R} + (1-\lambda)\mathbf{I}]^{-1}\right|^{-1/2} \times \exp\left\{-\frac{1}{2\sigma_v^2}(\boldsymbol{\theta} - X\beta)'[\lambda\mathbf{R} + (1-\lambda)\mathbf{I}](\boldsymbol{\theta} - X\beta)\right\}$;
- $[\sigma_v^2|\boldsymbol{\theta}, \beta, \lambda] \sim \mathrm{IG}\left[a_0 + \frac{m}{2}, b_0 + \frac{1}{2}(\boldsymbol{\theta} - X\beta)'\mathbf{D}(\boldsymbol{\theta} - X\beta)\right]$.

A.3. Gibbs sampling full conditional distributions under Model 3: YCM.

- $[\theta_i \mid y_i, \beta, \sigma_i^2, \sigma_v^2] \sim \mathrm{N}[\gamma_i y_i + (1-\gamma_i)\mathbf{x}_i'\beta, \sigma_i^2\gamma_i]$, where $\gamma_i = \sigma_v^2/(\sigma_v^2+\sigma_i^2)$, for $i = 1, ..., m$;
- $[\beta|\boldsymbol{\theta}, \sigma_v^2] \propto \mathrm{N}\left[\left(\sum_{i=1}^m \mathbf{x}_i\mathbf{x}_i'\right)^{-1}\left(\sum_{i=1}^m \mathbf{x}_i\theta_i\right), \sigma_v^2\left(\sum_{i=1}^m \mathbf{x}_i\mathbf{x}_i'\right)^{-1}\right]$;
- $[\sigma_i^2 \mid y_i, \theta_i] \sim \mathrm{IG}\left(a_i + \frac{d_i+1}{2}, b_i + \frac{(y_i-\theta_i)^2 + d_i s_i^2}{2}\right)$, where $d_i = n_i - 1$, for $i = 1, ..., m$;
- $[\sigma_v^2 \mid \boldsymbol{\theta}, \beta] \sim \mathrm{IG}\left[a_0 + \frac{1}{2}m, b_0 + \frac{1}{2}\sum_{i=1}^m (\theta_i - \mathbf{x}_i'\beta)^2\right]$.

A.4. Gibbs sampling full conditional distributions under Model 4: CAR-YCM.

- $[\boldsymbol{\theta} \mid \mathbf{y}, \beta, \lambda, \sigma_v^2, \sigma_i^2] \sim \mathrm{MVN}(\boldsymbol{\Lambda}\mathbf{y} + (\mathbf{I}-\boldsymbol{\Lambda})\mathbf{X}\beta, \boldsymbol{\Lambda}\mathbf{E})$, where $\boldsymbol{\Lambda} = (\mathbf{E}^{-1}+\mathbf{D}/\sigma^2)^{-1}\mathbf{E}^{-1}$, and $\mathbf{E} = \mathrm{diag}\{\sigma_1^2, ..., \sigma_m^2\}$, $\mathbf{D} = \lambda\mathbf{R} + (1-\lambda)\mathbf{I}$;
- $[\beta \mid \boldsymbol{\theta}, \lambda, \sigma_v^2] \sim \mathrm{MVN}[(\mathbf{X}'\mathbf{D}\mathbf{X})^{-1}\mathbf{X}'\mathbf{D}\boldsymbol{\theta}, \sigma_v^2(\mathbf{X}'\mathbf{D}\mathbf{X})^{-1}]$;
- $[\lambda \mid \boldsymbol{\theta}, \beta, \sigma_v^2] \propto \left|[\lambda\mathbf{R} + (1-\lambda)\mathbf{I}]^{-1}\right|^{-1/2} \times \exp\left\{-\frac{1}{2\sigma_v^2}(\boldsymbol{\theta} - X\beta)'[\lambda\mathbf{R} + (1-\lambda)\mathbf{I}](\boldsymbol{\theta} - X\beta)\right\}$

  $[\sigma_i^2 \mid y_i, \theta_i] \sim \mathrm{IG}\left(a_i + \frac{d_i+1}{2}, b_i + \frac{(y_i-\theta_i)^2 + d_i s_i^2}{2}\right)$, where $d_i = n_i - 1$, for $i = 1, ..., m$;
- $[\sigma_v^2 \mid \boldsymbol{\theta}, \beta, \lambda] \sim \mathrm{IG}\left[a_0 + \frac{m}{2}, b_0 + \frac{1}{2}(\boldsymbol{\theta} - X\beta)'\mathbf{D}(\boldsymbol{\theta} - X\beta)\right]$.

Appendix B

List of 20 health regions in the province of British Columbia with the corresponding sample sizes and spatial structures

ID number	Health region name	Sample size	Number of neighbours	Neighbours
1	East Kootenay	645	3	2, 3, 15
2	West Kootenay-Boundary	705	3	1, 3, 4
3	North Okanagan	890	5	1, 2, 4, 5, 15
4	South Okanagan Similameen	1,063	4	2, 3, 5, 6
5	Thompson	982	7	3, 4, 6, 9, 11, 12, 15
6	Fraser Valley	1,125	5	4, 5, 7, 8, 9
7	South Fraser Valley	1,437	4	6, 8, 17, 19
8	Simon Fraser	1,165	5	6, 7, 9, 17, 18
9	Coast Garibaldi	623	5	5, 6, 8, 11, 18
10	Central Vancouver Island	1,077	2	11, 20
11	Upper Island/Central Coast	746	4	5, 9, 10, 12
12	Cariboo	673	4	5, 11, 13, 15
13	North West	650	3	12, 14, 15
14	Peace Liard	611	2	13, 15
15	Northern Interior	859	6	1, 3, 5, 12, 13, 14
16	Vancouver	1,285	4	17, 18, 19, 20
17	Burnaby	871	5	7, 8, 16, 18, 19
18	North Shore	842	4	8, 9, 16, 17
19	Richmond	828	3	7, 16, 17
20	Capital	1,225	2	10, 16

Note that Vancouver (#16) and Capital (#20) are not adjacent regions in the map since they are separated by the ocean. However, due to the intensive and close connection between these two regions, we define them as neighbours in our study for illustration purpose only.

Appendix C

Direct and model-based point estimates and CVs

Comparison of point estimates					
Area ID	**Direct Est.**	**FHM**	**CAR-FHM**	**YCM**	**CAR-YCM**
1	0.0765	0.0793	0.0812	0.0795	0.0812
2	0.0804	0.0795	0.0793	0.0797	0.0794
3	0.0745	0.0726	0.0731	0.0725	0.0729
4	0.0893	0.0868	0.0874	0.0867	0.0873
5	0.0782	0.0739	0.0736	0.0729	0.0731
6	0.0943	0.0914	0.0927	0.0918	0.0928
7	0.0702	0.0707	0.0712	0.0711	0.0717
8	0.0858	0.0845	0.0848	0.0844	0.0849
9	0.0877	0.0763	0.0745	0.0765	0.0747
10	0.0763	0.0805	0.0799	0.0805	0.0796
11	0.0661	0.0685	0.0678	0.0679	0.0676
12	0.0717	0.0681	0.0681	0.0678	0.0677
13	0.0631	0.0687	0.0692	0.0690	0.0693
14	0.0673	0.0685	0.0680	0.0685	0.0686
15	0.0793	0.0721	0.0707	0.0728	0.0713
16	0.0657	0.0696	0.0702	0.0697	0.0704
17	0.0859	0.0778	0.0759	0.0773	0.0759
18	0.0583	0.0626	0.0633	0.0618	0.0626
19	0.0619	0.0649	0.0647	0.0653	0.0647
20	0.0877	0.0923	0.0914	0.0917	0.0908
Comparison of CVs					
Area ID	**Direct Est.**	**FHM**	**CAR-FHM**	**YCM**	**CAR-YCM**
1	0.168	0.107	0.099	0.107	0.100
2	0.127	0.105	0.104	0.097	0.093
3	0.135	0.116	0.106	0.110	0.097
4	0.102	0.084	0.076	0.079	0.072
5	0.158	0.094	0.076	0.105	0.083
6	0.113	0.086	0.080	0.086	0.081
7	0.124	0.099	0.096	0.106	0.101
8	0.102	0.085	0.076	0.081	0.073
9	0.158	0.119	0.105	0.117	0.105
10	0.121	0.087	0.086	0.086	0.084
11	0.141	0.118	0.108	0.109	0.105
12	0.196	0.119	0.109	0.130	0.116
13	0.168	0.115	0.108	0.111	0.108
14	0.206	0.126	0.125	0.136	0.133
15	0.121	0.101	0.087	0.094	0.083
16	0.127	0.101	0.097	0.103	0.097
17	0.124	0.107	0.100	0.105	0.096
18	0.155	0.143	0.136	0.134	0.130
19	0.154	0.135	0.134	0.128	0.128
20	0.103	0.086	0.085	0.083	0.082

References

Arora, V., and Lahiri, P. (1997). On the superiority of the Bayesian method over the BLUP in small area estimation problems. *Statistica Sinica*, 7, 1053-1063.

Bayarri, M.J., and Berger, J.O. (2000). P values for composite null models. *Journal of the American Statistical Association*, 95, 1127-1142.

Béland, Y. (2002). Canadian Community Health Survey Methodological Overview. Health Report, Statistics Canada, Catalogue no. 82-003, 13, 3, 9-14.

Besag, J., York, J. and Mollie, 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., Richardson, S. and Thomson, A. (2005). A comparison of Bayesian spatial models for disease mapping. *Statistical Methods in Medical Research*, 14, 35-39.

Brown, G., Chambers, R., Heady, P. and Heasman, D. (2001). Evaluation of small area estimation methods – An application to unemployment estimates from the UK LFS. Proceedings: Symposium 2001, *Achieving Data Quality in a Statistical Agency*: *A Methodological Perspective*, CD-ROM, 1-10.

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

Cressie, N. (1990). Small area prediction of undercount using the general linear model. Proceedings: Symposium 1990, *Measurement and Improving Data Quality*, Statistics Canada, 93-105.

Daniels, M.J., and Gatsonis, C. (1999). Hierarchical generalized linear models in the analysis of variations in health care utilization. *Journal of the American Statistical Association*, 94, 29-42.

Datta, G.S., Lahiri, P., Maiti, T. and Lu, K.L. (1999). Hierarchical Bayes estimation of unemployment rates for the states of the U.S. *Journal of the American Statistical Association*, 94, 1074-1082.

Dick, P. (1995). Modelling net undercoverage in the 1991 Canadian census. *Survey Methodology*, 21, 45-54.

Fay, R.E., and Herriot, R.A. (1979). Estimation of income for small places: An application of James-Stein procedures to census data. *Journal of the American Statistical Association*, 74, 268-277.

Gelfand, A.E. (1996). Model determination using sampling-based methods. In *Markov Monte Carlo in Practice* (Eds., W.R. Gilks, S. Richardson and D.J. Spiegelhalter), London: Chapman & Hall, 145-161.

Gelman, A., Carlin, J.B., Stern, H.S. and Rubin, D.B. (2004). *Bayesian Data Analysis*, 2nd Edition. Chapman & Hall/CRC.

Gelman, A., Meng, X.L. and Stern, H.S. (1996). Posterior predictive assessment of model fitness via realized discrepancies (with discussion). *Statistica Sinica*, 6, 733-807.

Gelman, A., and Rubin, D.B. (1992). Inference from iterative simulation using multiple sequences. *Statistical Science*, 7, 457-472.

Ghosh, M., Natarajan, K., Stroud, T.W.F. and Carlin, B.P. (1998). Generalized linear models for small area estimation. *Journal of the American Statistical Association*, 93, 273-282.

Ghosh, M., Natarajan, K., Walter, L.A. and Kim, D.H. (1999). Hierarchical Bayes GLMs for the analysis of spatial data: An application to disease mapping. *Journal of Statistical Planning and Inference*, 75, 305-318.

He, Z., and Sun, D. (2000). Hierarchical Bayes estimation of hunting success rates with spatial correlations. *Biometrics*, 56, 360-367.

Jiang, J., and Lahiri, P. (2006). Mixed model prediction and small area estimation (with discussion). *Test*, 15, 1-96.

Lele, S.R., Dennis, B. and Lutscher, F. (2007). Data cloning: Easy maximum likelihood estimation for complex ecological models using Bayesian Markov chain Monte Carlo methods. *Ecology Letters*, 10, 551-563.

Lele, S.R., Nadeem, K. and Schmuland, B. (2010). Estimability and likelihood inference for generalized linear mixed models using data cloning. *Journal of the American Statistical Association*, 105, 1617-1625.

Leroux, B.G., Lei, X. and Breslow, N. (1999). Estimation of disease rates in small areas: A new mixed model for spatial dependence. In *Statistical Models in Epidemiology, the Environment and Clinical Trials*, (Eds., M.E. Halloran and D. Berry). New York: Springer Verlag, 135-178.

Liu, B., Lahiri, P. and Kalton, G. (2008). Hieararchical Bayes modeling of survey weighted small area proportions. Unpublished Manuscript.

Maiti, T. (1998). Hierarchical Bayes estimation of mortality rates for disease mapping. *Journal of Statistical Planning and Inference*, 69, 339-348.

Mohadjer, L., Rao, J.N.K., Liu, B., Krenzke, T. and van de Kerckhove, W. (2007). Hierarchical Bayes small area estimates of adult literacy using unmatched sampling and linking models. *Proceedings of the American Statistical Association, Section of Survey Method Research.*

Moura, F.A.S., and Migon, H.S. (2002). Bayesian spatial models for small area proportions. *Statistical Modelling*, 2, 3, 183-201.

MacNab, Y.C. (2003). Hierarchial Bayesian spatial modeling of small-area rates of non-rare disease. *Statistics in Medicine*, 22, 1761-1773.

Maiti, T. (1998). Hierarchical Bayes estimation of mortality rates for disease mapping. *Journal of Statistical Planning and Inference*, 69, 339-348.

Meng, X.L. (1994). Posterior predictive p value. *The Annals of Statistics*, 22, 1142-1160.

Mollié, A. (1996). Bayesian mapping of disease. In *Markov Chain Monte Carlo in Practice*. London: Chapman and Hall, 359-379.

Rao, J.N.K. (2003). *Small Area Estimation*. New York: John Wiley & Sons, Inc.

Singh, A.C., Folsom, R.E., Jr. and Vaish, A.K. (2005). Small area modeling for survey data with smoothed error covariance structure via generalized design effects. Federal Committee on Statistical methods Conference proceedings, Washington, D.C., www.fcsm.gov.

Singh, B.B., Shukla, G.K. and Kundu, D. (2005). Spatio-temporal models in small area estimation. *Survey Methodology*, 31, 183-195.

Sinharay, S., and Stern, H.S. (2003). Posterior predictive model checking in hierarchical models. *Journal of Statistical Planning and Inference*, 111, 209-221.

Spiegelhalter, D.J., Best, N., Carlin, B.P. and van de Linde, A. (2002). Bayesain measures of model complexity and fit. *Journal of Royal Statistical Society*, B, 64, 583-639.

Souza, D.F., Moura, F.A.S. and Migon, H.S. (2009). Small area population prediction via hierarchical models. *Survey Methodology*, 35, 203-214.

Wang, J., and Fuller, W. A. (2003). The mean square error of small area predictors constructed with estimated area variances. *Journal of the American Statistical Association*, 98, 716-723.

You, Y. (2008a). An integrated modeling approach to unemployment rate estimation for sub-provincial areas of Canada. *Survey Methodology*, 34, 19-27.

You, Y. (2008b). Small area estimation using area level models with model checking and applications. *Proceedings of Survey Methods Section, Statistical Society of Canada.*

You, Y., and Chapman, B. (2006). Small area estimation using area level models and estimated sampling variances. *Survey Methodology*, 32, 97-103.

You, Y., and Rao, J.N.K. (2000). Hierarchical Bayes estimation of small area means using multi-level models. *Survey Methodology*, 26, 173-181.

You, Y., and Rao, J.N.K. (2002). Small area estimation using unmatched sampling and linking models. *The Canadian Journal of Statistics*, 20, 3-15.