Published in final edited form as:

Spat Stat. 2014 May 1; 8: 69–85. doi:10.1016/j.spasta.2013.12.001.

A comparison of spatial smoothing methods for small area estimation with sampling weights

Laina Mercer^a, Jon Wakefield^{a,*}, Cici Chen^b, and Thomas Lumley^c

^aDepartment of Statistics, University of Washington, United States ^bDepartment of Biostatistics, Brown University, United States ^cDepartment of Statistics, University of Auckland, New Zealand

Abstract

Small area estimation (SAE) is an important endeavor in many fields and is used for resource allocation by both public health and government organizations. Often, complex surveys are carried out within areas, in which case it is common for the data to consist only of the response of interest and an associated sampling weight, reflecting the design. While it is appealing to use spatial smoothing models, and many approaches have been suggested for this endeavor, it is rare for spatial models to incorporate the weighting scheme, leaving the analysis potentially subject to bias. To examine the properties of various approaches to estimation we carry out a simulation study, looking at bias due to both non-response and non-random sampling. We also carry out SAE of smoking prevalence in Washington State, at the zip code level, using data from the 2006 Behavioral Risk Factor Surveillance System. The computation times for the methods we compare are short, and all approaches are implemented in R using currently available packages.

Keywords

Complex surveys; Design-based inference; Intrinsic CAR models; Random effects models; Weighting

1. Introduction

Small area estimation (SAE) is used in many fields including education and epidemiology, and global, environmental and public health. Often the surveys carried out to inform SAE are complex in nature, with non-random sampling being carried out for reasons of necessity (i.e., logistical reasons) or to ensure that certain populations of interest are well represented. In addition, post-stratification may be used to reweight the observations in order to recover known population totals. This approach can account for non-response within the strata used in the post-stratification.

There are two approaches to modeling complex survey data that we shall consider in this paper. In the first *design-based* approach weighted estimators are considered, with inference

carried out based on the (randomization) distribution of the samples that could have been collected, i.e., the distribution of the individuals that could appear in the sample. In contrast, a *model-based* approach assumes a hypothetical infinite population from which the responses are drawn. While appealing from a conceptual point of view (since standard statistical modeling machinery can be leaned upon), the modeling approach is difficult to implement since one must model the sampling mechanism, if informative, at least to some extent. For example, if non-random sampling is based on particular inclusion variables (e.g., race or geographical area) then these variables must be included in the model if they are associated with the outcome of interest. Similarly, variables that affect the probabilities of non-response must also be included in the model, again if they are related to the outcome. The alternative is to assume that variables upon which sampling is based and non-response depends are unrelated to the outcome of interest, which is a dangerous endeavor. Another impediment to the model-based approach is that the key variables that are required for inclusion may be unavailable in public-use databases. Even if available, the sampling scheme may be highly complex, requiring a model which has a large number of parameters and being therefore difficult to fit. Gelman (2007b) describes the issues, and the accompanying discussion (Bell and Cohen, 2007; Breidt and Opsomer, 2007; Little, 2007; Lohr, 2007; Pfefferman, 2007; Gelman, 2007a) gives a range of perspectives on the use of weighted estimators, regression modeling, or a combination of the two.

In this paper we will consider SAE in the situation in which either the variables upon which sampling was based are unavailable or the scheme is so complex that a simpler approach is desired. SAE has seen a great deal of research interest, with Rao (2003) being a classic text. In the related field of disease mapping, the use of spatial modeling is commonplace (Wakefield et al., 2000), but in this context the data usually consist of a complete enumeration of disease cases in an area, so that no weighting scheme needs to be considered. It is the existence of the weights that causes a major difficulty when one wishes to use spatial smoothing in SAE, and consequently there are relatively few instances of approaches that use spatial smoothing within a model that acknowledges the sampling scheme. In Chen et al. (submitted for publication) a new method of incorporating the weights within a spatial hierarchical model was introduced, and various random effects models were compared via simulation. In this paper we compare the method with a number of other suggested methods for weighting.

As a motivating example, we examine data from the Behavioral Risk Factor Surveillance System (BRFSS). This survey is carried out at the state level in the United States and is the largest telephone-based survey in the world. In the BRFSS survey, interviewees (who are 18 years or older) are asked a series of questions on their health behaviors and provide general demographic information, such as age, race, gender and the zip code in which they live. In this paper we focus on the survey conducted in Washington State in 2006, and on the Centers for Disease Control (CDC) calculated variable *Adults who are current smokers*. With respect to this question, 19,502 respond with “No”, 3733 with “Yes” and 132 were classified as “don’t know/refuse/missing”. In the analysis, we remove these latter values. The response variable is therefore a binary indicator and our objective is to estimate the number of individuals who are 18 or older and who are current smokers, in each of 498 zip

codes in Washington State. We also utilize population estimates from 2006. Table 1 summarizes the population and survey data. So far as the survey is concerned, the number of samples per zip code shows large variability with a median of 30 and minimum and maximum values of 1 and 384. The spread is apparent in Fig. 1. Fig. 2 maps, by zip code, the observed number of smokers in the sample (top) and the sample sizes (bottom) and the spatial variability in each map is evident.

We now describe in greater detail the complex survey scheme that was used by BRFSS in 2006. In this year, the BRFSS survey used land-lines only, and utilized a disproportionate stratified random sample scheme with stratification by county and “phone likelihood”. Under this scheme in each county, based on previous surveys, blocks of 100 telephone numbers were classified into strata that are either “likely” or “unlikely” to yield residential numbers. Telephone numbers in the “likely” strata are then sampled at a higher rate than their “unlikely” counterparts. Once a person is reached at a phone number the number of eligible adults (aged 18 or over) is determined, and one of these is randomly selected for interview. The sample weight, Sample Wt, is then calculated as the product of four terms

$$\text{Sample Wt} = \text{Strat Wt} \times \frac{1}{\text{No Telephones}} \times \text{No Adults} \times \text{Post Strat Wt} \quad (1)$$

where Strat Wt is the inverse probability of a “likely” or “unlikely” stratum being selected in a particular county, No Telephones represents the number of residential telephones in the respondent’s household, No Adults is the number of adults in the household, and Post Strat Wt is the posts-tratification correction factor. The latter is given by the number of people in strata defined by gender and age, using the 7 age groups 18–24, 25–34, 35–44, 45–54, 55–64, 65–74, 75+. The raw data we will base estimation on are the respondent’s outcome, with an accompanying weight, and the population information. And crucially, we will also examine the possibility of leveraging geographic information to smooth rates across zip codes.

The structure of the paper is as follows. In Section 2 we describe a number of approaches to formulating hierarchical models that incorporate weighting and in Section 3 a number of these methods are compared via a simulation study. In Section 4 we return to the BRFSS data and the paper concludes with a discussion in Section 5.

2. Methods

2.1. Notation and the Horvitz–Thompson estimator

We first establish our notation. We will focus on binary outcomes, and let Y_{ik} represent the binary indicator for the event of interest on the k -th individual, $k = 1, \dots, N_i$ in the i th area, $i = 1, \dots, I$. Common small area characteristics of interest include the true total count,

$T_i = \sum_{k=1}^{N_i} Y_{ik}$, or the true proportion, $P_i = \frac{T_i}{N_i}$, in area i , $i = 1, \dots, I$. In common with the majority of the survey sampling literature we will denote population values with upper case letters and sampled values with lower case letters. To obtain estimates, a survey is conducted with probabilities of being sampled for the k -th person in area i being denoted π_{ik} .

We use s_i to indicate the set of individuals who are sampled from area i with y_{ik} being the observed value for $k \in s_i$ with $|s_i| = m_i$, so that the latter is the sample size in area i . The design weights are calculated as the reciprocal of the sampling probabilities for selection, so that $w_{ik}^d = \pi_{ik}^{-1}$.

We will focus on estimation of the total T_i . A common and famous estimator is that introduced by Horvitz and Thompson (1952). The Horvitz–Thompson estimator is:

$$\hat{T}_i = \sum_{k \in s_i} \frac{y_{ik}}{\pi_{ik}}. \quad (2)$$

The estimated design variance of the estimator (2) of T_i , over the randomization distribution (i.e. over the distribution of all samples of size m_i that could have been selected in area i) is

$$v\hat{ar}(\hat{T}_i) = \sum_{k \in s_i, k' \in s_i} \left(\frac{y_{ik} y_{ik'}}{\pi_{ik} \pi_{ik'}} - \frac{y_{ik} y_{ik'}}{\pi_{ikk'}} \right) \quad (3)$$

where $\pi_{ikk'}$ is the sampling probability for the pair of individuals k and k' in area i . A common strategy is to produce sample weights as the product of the design weights and the post-stratification weights w_{ik}^p :

$$w_{ik} = w_{ik}^d \times w_{ik}^p. \quad (4)$$

For a description of the construction of post-stratification weights, see Lumley (2010, Section 7.2). We have already seen an example of this construction as Eq. (1), in the context of the BRFSS example. When post-stratification is carried out the Horvitz–Thompson estimator is

$$\hat{T}_i = \sum_{k \in s_i} w_{ik} y_{ik} \quad (5)$$

and the variance formula is more complex, see Chen et al. (submitted for publication) for details.

2.2. Sampling models

In this section, in anticipation of the development of a hierarchical smoothing model, we consider various approaches to constructing a likelihood for the observed data.

The simplest approach is to ignore the design and take

$$y_i | P_i \sim \text{Binomial}(m_i, P_i), \quad (6)$$

where $y_i = \sum_{k \in s_i} y_{ik}$. If the design is informative we would expect bias in the estimator y_i/m_i of P_i . Ignoring the weighting and using this model is often carried out in SAE, for examples see Rao (2003).

The next approach summarizes the data in area i via the asymptotic distribution of the estimator (2) (or the post-stratified version (5)) or, equivalently, via the asymptotic estimator of P_i , which we denote $\hat{P}_i = \hat{T}_i / \hat{N}_i$ with variance estimator $\text{var}(\hat{P}_i) = \text{var}(\hat{T}_i) / \hat{N}_i^2$. In this way the design is acknowledged in both the estimator and its variance. We could simply take $\hat{P}_i | P_i \sim N(P_i, \text{var}(\hat{P}_i))$, but this does not constrain the probability to lie in $(0, 1)$ which we might anticipate would cause difficulties, in particular in areas with small m_i . As an alternative we define the area-level data summary as $y_i^L = \log[\hat{P}_i / (1 - \hat{P}_i)]$ as the empirical logistic transform of $\hat{P}_i$. The likelihood is then taken as the asymptotic distribution

$$y_i^L | P_i \sim N \left(\log \left[\frac{P_i}{1 - P_i} \right], \frac{\text{var}(\hat{P}_i)}{\hat{P}_i^2 (1 - \hat{P}_i)^2} \right). \quad (7)$$

A number of authors have suggested an approach in which the likelihood is weighted. This approach is often referred to as *pseudo-likelihood*, to acknowledge the sampling design; early references are Binder (1983) and Skinner (1989). In the version we implement the weights are scaled as

$$w_{ik}^s = m_i \times \frac{w_{ik}}{\sum_{k=1}^{m_i} w_{ik}},$$

as recommended by Pfefferman et al. (1998), see also Asparouhov (2006). Defining

$y_i^P = \sum_{k \in s_i} w_{ik}^s y_{ik}$, the likelihood is taken as

$$y_i^P | P_i \sim \text{Binomial}(m_i, P_i). \quad (8)$$

The pseudo-likelihood approach has been used with a spatial smoothing model by Congdon and Lloyd (2010), with the weights being scaled to sum to the sample size m_i . These authors estimate diabetes prevalence for ZIP Code Tabulation Areas (ZCTAs) in the USA. A drawback with the general approach is that the appropriate standard error is not recovered in the case of clustering. Rabe-Hesketh and Skrondal (2006) utilize a pseudo-likelihood with scaled weights, and use sandwich estimation to provide valid standard error estimates. These authors embed this approach within a multilevel framework but do not consider spatial smoothing.

In a far-reaching paper, Raghunathan et al. (2007) describe an approach for combining data from different surveys. In this paper we consider their approach as applied to a single survey. They utilize the arcsin square root (also known as the angular) transform:

$y_i^A = \arcsin \left(\sqrt{\hat{P}_i} \right)$ which is the approximate variance stabilizing transformation for binary data and results in $-\pi/2 < y_i^A < \pi/2$. This, and closely related, transforms are discussed in Anscombe (1948). The “effective sample size” m_i^E is obtained by solving $\hat{P}_i(1 - \hat{P}_i) / m_i^E = \text{var}(\hat{P}_i)$ to give

$$m_i^E = \frac{\hat{P}_i (1 - \hat{P}_i)}{v \hat{a}r(\hat{P}_i)}. \quad (9)$$

The asymptotic likelihood used by Raghunathan et al. (2007) is

$$y_i^A | P_i \sim N \left(\arcsin \left(\sqrt{P_i} \right), \frac{1}{4m_i^E} \right). \quad (10)$$

Raghunathan et al. (2007) use (10) as the first stage within a hierarchical model, but do not introduce spatial random effects. The scaling of the weights with respect to the effective size has been considered by a number of authors including Potthoff et al. (1992) and Longford (1996), with the latter explicitly considering variance components models, though again without considering spatial smoothing.

A technical but important detail is that the above approach runs into difficulties when $\hat{P}_i = 0/1$. In these cases we use a procedure described in Chen et al. (submitted for publication) to produce an effective sample size.

Briefly, when $\hat{P}_i = 0$ we augment $\{y_{ik}, k \in s_i\}$ with an extra sampled individual with $y_{i,m_{i+1}} = 1$ and associated weight

$$w_{i,m_{i+1}} = \frac{\tilde{P}_i}{1 - \tilde{P}_i} \times \sum_{k \in s_i} w_{ik}$$

where $\tilde{P}_i$ is an empirical Bayes smoothed estimator for area i . In this way we obtain the estimator

$$\hat{P}_i = \frac{\sum_{k \in s_i} w_{ik} y_{ik} + w_{i,m_{i+1}}}{\sum_{k \in s_i} w_{ik} + w_{i,m_{i+1}}} = \frac{w_{i,m_{i+1}}}{\sum_{k \in s_i} w_{ik} + w_{i,m_{i+1}}} = \tilde{P}_i,$$

as we desire. An equivalent approach is available when $\hat{P}_i = 1$. This procedure is also used with the logistic model as given by (7).

Chen et al. (submitted for publication) build on this approach by defining the “effective number of cases” y_i^E as the product of the effective sample size m_i^E and the weighted proportion $\hat{P}_i$ to give

$$y_i^E = m_i^E \times \hat{P}_i. \quad (11)$$

The likelihood they assume is

$$y_i^E | P_i \sim \text{Binomial}(m_i^E, P_i).$$

The rationale here is that both numerator and denominator are adjusted for the sampling design, and the use of a binomial likelihood, though not the “true” likelihood, will better reflect the sampling distribution than a normal approximation. In Chen et al. (submitted for publication) different random effects models were compared, using the proposed method and direct estimation.

2.3. Hierarchical models

We examine the use of three-stage models with the first stage given by one of the forms, (6), (7), (8), (10), (11) described in the previous sections. At the second stage of the model we introduce the random effects on the transformed scale and denote the area-specific parameter on this scale as θ_i . For model (10) this is the arcsin square root scale, while for all other models it is the logistic scale. We consider two different second stages for each of the five likelihoods, independent normal random effects only, and independent plus spatial random effects. The non-spatial normal second stage is defined as

$$\theta_i = \beta_0 + V_i, \quad (12)$$

with $V_i | \sigma_v^2 \sim \text{i.i.d. } N(0, \sigma_v^2)$, so that for four of the models $\exp(\beta_0)$ is the area-level odds of the event of interest in an area with $V_i = 0$. Another interpretation is as the median odds of the event of interest across areas. For model (10), β_0 is the arcsin square root of the frequency of the event of interest in an area with $V_i = 0$.

The second random effects model we consider is the “convolution” model (Besag et al., 1991):

$$\theta_i = \beta_0 + V_i + U_i, \quad (13)$$

with $V_i | \sigma_v^2 \sim \text{i.i.d. } N(0, \sigma_v^2)$ and U_i following an intrinsic conditional autoregressive (ICAR) model (Besag and Kooperberg, 1995; Rue and Held, 2005). The ICAR model is a non-parametric, stochastic smoothing model with

$$U_i | U_j, j \in \text{ne}(i), \sigma_u^2 \sim N \left(\bar{U}_i, \frac{\sigma_u^2}{n_i} \right), \quad (14)$$

where $\text{ne}(i)$ indexes the set of neighbors of area i , n_i is the number of such neighbors and

$\bar{U}_i = \frac{1}{n_i} \sum_{j \in \text{ne}(i)} U_j$ is the mean of the neighbors. In what follows we take the conventional approach of spatial epidemiology in which two areas are considered neighbors if they share a common boundary. Hence, we see that in (14) the spatial random effect is shrunk towards the mean of its neighbors, with the shrinkage being more pronounced for areas with more neighbors. To ensure identifiability in a model with an intercept a sum-to-zero constraint is placed on $U_1, \dots, U_I$ (Besag and Kooperberg, 1995), which is equivalent to constraining the collection $U_1, \dots, U_I$ to have zero mean.

In general one may specify proper subjective priors for σ_v^2 and σ_u^2 based on the context. However, when one carries out a simulation study to evaluate Bayesian procedures, one is faced with the thorny question of which priors to use, so that one does not favor one approach over another. In our case this is pertinent since the two scales have quite different ranges on the random effects, for example, the whole of the real line for the logit models, and $[-\pi/2, \pi/2]$ for the arcsin square root model.

Browne and Draper (2006a) carried out an extensive simulation study in which the bias and coverage probabilities were examined as a function of various characteristics, including the priors on the variance components. They found that the uniform prior $U(0, 1/\varepsilon)$ (for small ε), or an improper uniform prior, on a generic random effects σ^2 produced reasonable behavior. Lambert (2006) prefers uniform priors on the standard deviation, σ , which is further supported by Gelman (2006) who states: “In fitting hierarchical models, we recommend starting with a noninformative uniform prior density on standard deviation parameters σ ”. This view is also supported by Browne and Draper (2006b). In the simulation study of Section 3 we take an improper uniform prior on σ_v and, to aid in stability, a Gamma(0.5, 0.008) prior on the spatial conditional precision σ_u^{-2} . The latter prior gives a 95% range on the more interpretable σ_u scale of (0.056, 4.04). We use a normal prior with large variance for β_0 .

2.4. Inference for counts

The point estimate of the population count of interest T_i is

$$\hat{T}_i = \hat{P}_i \times N_i, \quad (15)$$

where $\hat{P}_i$ is the direct estimator (5) and the variance is

$$\hat{\text{var}}(\hat{T}_i) = \hat{\text{var}}(\hat{P}_i) \times N_i^2. \quad (16)$$

Under a Bayesian approach one may summarize the posterior distribution for T_i using quantiles. If a point estimate is required then it is given by (15) with $\hat{P}_i$ being replaced by the posterior mean or median. The posterior variance $\text{var}(T_i|y)$ is given by (16) with $\hat{\text{var}}(\hat{P}_i)$ replaced by the posterior variance $\text{var}(P_i|y)$.

2.5. Implementation

The hierarchical models described above may be implemented within the R programming environment using the survey package (Lumley, 2010) to obtain the appropriate variances (and hence effective sample sizes). To fit the random effects models we use the integrated nested Laplace approximation (INLA) approach (Rue et al., 2009), which also has an R implementation. INLA is very fast and applies numerical integration to the fixed effects (β_0 , σ_v , σ_u) and analytic (Laplace) integral approximations to the random effects (V_i , U_i) (nested within levels of (β_0 , σ_v , σ_u)).

3. Simulation study

We now present a simulation study to compare five of the estimators described in the previous section. The estimators we compare are the naive binomial (6), the logit normal (7), pseudo-likelihood (8), the arcsin square root transform (10) and the numerator and denominator effective sample size adjusted binomial (11). In each case we consider two random effects models: independent random effects only, and the convolution model with both independent and spatial ICAR random effects. We also include a pair of models with no hierarchical smoothing. One uses the binomial model (6) and the other uses the Horvitz–Thompson estimator and its associated variance estimator (and therefore adjusts for design bias). We follow Chen et al. (submitted for publication) and report two sets of simulations, one to address non-response bias, and another to address selection bias.

To evaluate the estimates, three statistics will be compared: the squared bias, the variance and the mean squared error (MSE). Let S denote the total number of simulations (which we take as $S = 100$) and T_i the true (but unobserved) count of the event of interest in area i (which is kept constant across simulations). The summary statistics are

$$\begin{aligned} Bias &= \frac{1}{I} \sum_{i=1}^I (\bar{\hat{T}}_i - T_i), \text{ where } \bar{\hat{T}}_i = \frac{1}{S} \sum_{s=1}^S \hat{T}_i^{(s)}, \\ Variance &= \frac{1}{I} \sum_{i=1}^I \left(\frac{1}{S-1} \sum_{s=1}^S (\hat{T}_i^{(s)} - \bar{\hat{T}}_i)^2 \right), \\ MSE &= Bias^2 + Variance. \end{aligned}$$

Good methods display low MSE.

3.1. Non-response bias

In all simulation studies, we take as geography the zip codes of Washington State. For direct comparison with the results in Chen et al. (submitted for publication) we set the parameters of the simulation based on diabetes. We simulate cases using a probability of diabetes p_{ij} for individuals in area i and post-stratification group j . There are $J = 6$ groups consisting of three age bands and two genders. We examine five scenarios with varying prevalence and response rates. In each of the five scenarios, simple random sampling of individuals is carried out within each area. However, individuals within area i respond to the survey within post-stratification group j with response probabilities q_{ij} , $i = 1, \dots, I$, $j = 1, \dots, J$. Thus, we assume that missingness depends on post-stratification group only and so, conditional on the post-stratification group, the missingness does not depend on the observed response. The sample sizes m_i are taken as the actual number of individuals who responded in the Washington State 2006 BRFSS survey.

Scenario 1—In scenario 1 we consider the ideal situation in which every individual selected responds to the survey. The prevalence of diabetes in area i and group j , are the same in each area so that $p_{ij} = p_j$; these values are given in Table 2.

Scenario 2—In scenario 2 we introduce non-response with the response rate being the same in each area but differing by age and gender, i.e. $q_{ij} = q_j$ for $j = 1, \dots, 6$. The rates used

are given in Table 3. The response rates increase with age, and women have higher rates than men.

Scenario 3—In this scenario the response rates for each group vary between areas via the stochastic relationship:

$$\text{logit}(q_{ij}) = \text{logit}(q_j) + b \times \varepsilon_i,$$

where $\varepsilon_i \sim \text{i.i.d. } N(0, 1)$. The median response rate, $\exp(q_j)/[1 + \exp(q_j)]$ is the same as in Scenario 2. We set $b = 0.35$ to give 95% ranges for the response rates in each of the six groups as given in parentheses in Table 3.

Scenario 4—In scenario 4 we introduce spatial dependence into the underlying prevalence rates. This dependency is induced by adding a spatially correlated area-level covariate x_i :

$$\text{logit}(p_{ij}) = \text{logit}(p_j) + b \times x_i.$$

To simulate spatially correlated covariates x_i , we employ a zero mean, unit variance ICAR model. Details on how to simulate from ICAR models can be found in Rue and Held (2005). We choose $b = 0.2$ to allow variation in the prevalence rates between area; Table 2 gives the marginal (across areas) 95% ranges for the prevalence rates. In this scenario everyone responds.

Scenario 5—In scenario 5 we allow the response rate for each group to vary between areas by adding a spatial component to the variation:

$$\text{logit}(q_{ij}) = \text{logit}(q_j) + b \times x_i,$$

where x_i is again simulated from a zero mean, unit variance ICAR model. We set $b = 0.3$ to give the 95% ranges in Table 3.

The results of this simulation are summarized in Table 4 and we make the following observations:

- The non-hierarchical models have the lowest bias since there is no shrinkage. In scenarios 1 and 4 in which there is full response the unadjusted estimator has smallest bias while for all other scenarios the Horvitz–Thompson estimator is best. For the hierarchical models, the spatial models have lower bias than the independent models.
- The variance of the non-hierarchical models is very large, clearly showing the benefits of hierarchical modeling, as is well-known. The unadjusted binomial, logit and pseudo-likelihood models have relatively low variance with the independent version of the last of these giving the lowest variance.
- In terms of MSE the effective sample size spatial model gives the best performance in the three scenarios in which there is non-response bias (scenarios 2, 3 and 5),

closely followed by the pseudo-likelihood spatial model, with the arcsin square root having the next best performance. In scenarios 1 and 4 (which recall have full response) the spatial unadjusted binomial model slightly out-performs the effective sample size model.

3.2. Selection bias

To investigate the potential for selection bias, we let Z_{ik} denote a binary design variable that dictates whether the k -th individual in area i is selected to be surveyed or not. We use the population simulated from scenario 1 in the simulation study for non-response, and assign the status of the design variable for individual k in area i based on

$$\Pr(Z_{ik}=1|Y_{ik}=1)=s, \quad \Pr(Z_{ik}=1|Y_{ik}=0)=0.1.$$

These probabilities provide a correlation between the design variable Z and the outcome variable Y and we examine the extent of the correlation by assigning s values of 0.1, 0.3, 0.5 and 0.8. We still take the zip code geography of Washington State and the total sample size m of the 2006 Washington State BRFSS. Let P_i^z denote the proportion of population with $Z = 1$ in area i . We set the sample size $m_i = mp_i^z / \sum p_i^z$, and within each area $m_i/2$ are selected with $Z = 1$ and $m_i/2$ with $Z = 0$. Oversampling populations with certain characteristics is a common technique in surveys. The information on the variable Z is only used when conducting the survey (and in calculating the sample weights), and is considered unavailable at the time of analysis.

Table 5 gives the results, and the picture is not as clear cut as with the non-response set of simulations. However, we make the following observations:

- In terms of bias, again the non-hierarchical approaches are best since they do not employ shrinkage. The adjusted estimators have the lowest bias in all cases but the one in which there is no selection bias ($s = 0.1$), in which case the unadjusted estimator performs best.
- With respect to variance, the independent random effects models perform best with the unadjusted binomial having the lowest variance in all cases but the one with the worst selection bias. For this case the independent pseudo-likelihood model has the lowest variance.
- In terms of MSE, if there is no selection bias the unadjusted hierarchical models both perform well. For the second and third levels of selection bias ($s = 0.3, 0.5$) the pseudo-likelihood approaches perform best, with the spatial versions being a little better than the independent version. For the most extreme selection bias case each of the adjusted hierarchical models perform reasonably, with the effective sample size model being best.

In conclusion, no one model is superior in all situations, though hierarchical smoothing is clearly a good idea.

4. BRFSS example

We apply the sample weighted Bayesian hierarchical models we described in Section 2 to the Washington State 2006 BRFSS data introduced in Section 1. Sampling weights are taken to be the final weights used in the BRFSS survey, as in (1). These weights range between 1.2 and 4675 across zip codes. The effective sample sizes and number of observations used in the effective sample size approach are calculated using the design-based Horvitz–Thompson variance estimator. Fig. 3 gives the effective sample sizes, as calculated from (9), plotted against the observed sample sizes. In the majority of cases the effective sample size is lower than the observed sample size, so that the design is resulting in a loss in information, when compared to simple random sampling, in those areas.

We fitted each of the models that were considered in the simulations. With respect to priors, a $\text{Ga}(0.5, 0.0008)$ prior was initially taken for the spatial precision, σ_u^{-2} , with an improper uniform prior on the non-spatial standard deviation, σ_v . Later we examine sensitivity to the prior on σ_u^{-2} .

Fig. 4 presents the boxplots of logit-transformed estimates of adult smoking prevalence by zip code under different approaches. For comparison we also include the design-based estimates, which are denoted as “Direct” in the figure. It is clear that the direct estimates exhibit a large amount of between zip code variation, with some extreme values. The variation is significantly reduced by all of the Bayesian hierarchical models. The pseudo-likelihood approach and the effective sample size approach give very similar estimated adult smoking prevalences both in terms of the location and spread. The boxplots for the unadjusted binomial model have a slightly lower location, reflecting selection and non-response bias.

As we saw in Section 3, choosing an appropriate hierarchical model is not straightforward, with one possibility being to report not a single set of estimates. It is also interesting to identify areas with a large number of samples (so that a good idea of the “truth” may be obtained). Within these areas one may repeatedly select small samples (of size n) and then investigate, via MSE, which model can most accurately reproduce the totals. This procedure is carried out for three areas, zip codes 98801 ($m_i = 317$), 98802 ($m_i = 380$) and 99347 ($m_i = 331$). The results appear in Tables 6-8, with Table 9 summarizing over all three areas. Unfortunately the conclusions are far from clear cut for these data, though hierarchical modeling is obviously preferable. In zip code 98801 the use of the weights is clearly beneficial and the logit normal independent random effects model produces the smallest MSE. For zip code 98802 the use of the weights is not as beneficial and the unadjusted hierarchical spatial binomial model performs best. Finally, for zip code 99347, the logit normal spatial model gives the lowest MSE. Overall (Table 9) the logit normal spatial and unadjusted independent binomial models produce the lowest MSE. In Fig. 5 we see that, across the study region, the estimated smoking prevalence and total counts predicted by the logit normal spatial and unadjusted binomial model are quite similar. Here we report results under the logit normal spatial model.

In Fig. 6 we display a map of the estimated total number of adult smokers by zip code. The predicted counts are highest around the Puget Sound area (the channel running north–south with many small, highly populated, zip codes) and the central/south area. These areas correspond to King, Snohomish and Spokane counties and the Yakima valley, which are the most populated counties in Washington State. Fig. 7 provides a map of a measure of the uncertainty, namely the 95% intervals of the predicted smoking counts, using the spatial logit model. We see that, not surprisingly, the greatest uncertainty lies in the areas with the largest estimated counts.

To investigate the sensitivity of our estimates to the prior distribution selected for the spatial precision, σ_u^{-2} , we vary the prior and compare the posterior medians of σ_u and σ_v as well as the proportion of total variance contributed by the spatial component. In addition to the $\text{Ga}(0.5, 0.008)$ prior that was used for the simulation study we considered $\text{Ga}(1.0, 0.026)$ and $\text{Ga}(0.35, 0.001)$, which have a 95% range for the residual odds of (0.5, 2.0) and (0.1, 10), respectively (Wakefield, 2009). In Table 10 we report the variance parameter estimates and we see they are insensitive to the prior chosen for σ_u^2 . Maps of count estimates (not shown) show only small changes under the different priors. With 498 zip codes the insensitivity is not unexpected. In general, it is worth investigating the sensitivity of results to variance parameter prior specification, since we would not expect the stability seen here to be repeated when the number of areas is not so large. The values in Table 10 do vary by model, however, with the greatest discrepancy being between the arcsin square root scale and the other, logit, models. It is a little surprising that the proportion spatial is so much lower under the arcsin square root model, when compared to the models on the logit scale.

5. Discussion

In this paper we have considered random effects models that account for the sampling weights that are common in SAE. The simulations of Section 3 clearly illustrate the benefits of hierarchical modeling, namely large reductions in the variance of parameter estimation when compared with non-hierarchical approaches. These simulations also show that non-response and selection bias can be reduced via the incorporation of the weights. Further simulations are required to characterize situations in which different approaches may be advantageous, since we saw the choice of an optimal model was not clear in our application to BRFSS smoking data.

To implement the methods compared in this paper we have utilized the survey and inla packages within the R computing environment. The code to fit the models described in the paper is available at <http://faculty.washington.edu/jonno/software.html>. The INLA approximation is extremely fast and the method has been investigated in many different scenarios with Fong et al. (2010) looking specifically at generalized linear mixed models. They illustrate that the method is accurate for the situations considered, except for the case of sparse binomial data.

Acknowledgments

The first author was supported by a seed grant from the Center for Statistics and the Social Sciences. The second author was supported by grant R01 AI029168 from the National Institutes of Health.

References

- Anscombe F. The transformation of Poisson, binomial and negative-binomial data. *Biometrika*. 1948; 35:246–254.
- Asparouhov T. General multi-level modeling with sampling weights. *Comm Statist Theory Methods*. 2006; 35:439–460.
- Bell R, Cohen M. Comment on “Struggles with survey weighting and regression modeling”. *Statist Sci*. 2007; 22:165–167.
- Besag J, Kooperberg C. On conditional and intrinsic auto-regressions. *Biometrika*. 1995; 82:733–746.
- Besag J, York J, Mollié A. Bayesian image restoration with two applications in spatial statistics. *Ann Inst Statist Math*. 1991; 43:1–59.
- Binder D. On the variances of asymptotically normal estimators from complex surveys. *Internat Statist Rev*. 1983; 51:279–292.
- Breidt F, Opsomer J. Comment on “Struggles with survey weighting and regression modeling”. *Statist Sci*. 2007; 22:168–170.
- Browne W, Draper D. A comparison of Bayesian and likelihood-based methods for fitting multilevel models. *Bayesian Anal*. 2006a; 1:473–514.
- Browne W, Draper D. A comparison of Bayesian and likelihood-based methods for fitting multilevel models (rejoinder). *Bayesian Anal*. 2006b; 1:547–550.
- Chen C, Wakefield J, Lumley T. The use of sample weights in Bayesian hierarchical models for small area estimation. 2013 submitted for publication.
- Congdon P, Lloyd P. Estimating small area diabetes prevalence in the US using the behavioral risk factor surveillance system. *J Data Sci*. 2010; 8:235–252.
- Fong Y, Rue H, Wakefield J. Bayesian inference for generalized linear mixed models. *Biostatistics*. 2010; 11:397–412. [PubMed: 19966070]
- Gelman A. Prior distributions for variance parameters in hierarchical models. *Bayesian Anal*. 2006; 1:515–534.
- Gelman A. Rejoinder to “Struggles with survey weighting and regression modeling”. *Statist Sci*. 2007a; 22:184–188.
- Gelman A. Struggles with survey weighting and regression modeling. *Statist Sci*. 2007b; 22:153–164.
- Horvitz D, Thompson D. A generalization of sampling without replacement from a finite universe. *J Amer Statist Assoc*. 1952; 47:663–685.
- Lambert P. Comment on article by Browne and Draper. *Bayesian Anal*. 2006; 1:543–546.
- Little R. Comment on “Struggles with survey weighting and regression modeling”. *Statist Sci*. 2007; 22:171–174.
- Lohr S. Comment on “Struggles with survey weighting and regression modeling”. *Statist Sci*. 2007; 22:175–178.
- Longford N. Model-based variance estimation in surveys with stratified clustered design. *Aust J Stat*. 1996; 38:333–352.
- Lumley, T. *Complex Surveys: A Guide to Analysis Using R*. John Wiley and Sons; Hoboken, Jersey: 2010.
- Pfefferman D. Comment on “Struggles with survey weighting and regression modeling”. *Statist Sci*. 2007; 22:179–183.
- Pfefferman D, Skinner C, Holmes D, Goldstein H, Rasbash J. Weighting for unequal selection probabilities in multilevel models. *J R Stat Soc Ser B*. 1998; 60:23–40.
- Pothoff R, Woodbury M, Manton K. “Equivalent sample size” and “equivalent degrees of freedom” refinements for inference using survey weights under superpopulation models. *J Amer Statist Assoc*. 1992; 87:383–396.
- Rabe-Hesketh S, Skrondal A. Multilevel modelling of complex survey data. *J R Stat Soc Ser A*. 2006; 169:805–827.
- Raghunathan T, Xie D, Schenker N, Parsons V, Davis W, Dood K, Feuer E. Combining information from two surveys to estimate county-level prevalence rates of cancer risk factors and screening. *J Amer Statist Assoc*. 2007; 102:474–486.

- Rao, J. Small Area Estimation. John Wiley; New York: 2003.
- Rue, H.; Held, L. Gaussian Markov Random Fields: Theory and Application. Chapman and Hall/CRC Press; Boca Raton: 2005.
- Rue H, Martino S, Chopin N. Approximate Bayesian inference for latent Gaussian models using integrated nested Laplace approximations (with discussion). *J R Stat Soc Ser B*. 2009; 71:319–392.
- Skinner, C. Domain means, regression and multivariate analysis. In: Skinner, C.; Holt, D.; Smith, T., editors. *Analysis of Complex Surveys*. Wiley; Chichester: 1989. p. 59-87.
- Wakefield J. Multi-level modelling, the ecologic fallacy, and hybrid study designs. *Int J Epidemiol*. 2009; 38:330–336. [PubMed: 19339258]
- Wakefield, JC.; Best, NG.; Waller, LA. Bayesian approaches to disease mapping. In: Elliott, P.; Wakefield, JC.; Best, NG.; Briggs, D., editors. *Spatial Epidemiology: Methods and Applications*. Oxford University Press; Oxford: 2000. p. 104-127.

Fig. 1.
For 2006 Washington BRFSS data: histograms of *actual* sample sizes by zip code.

Fig. 2. Maps of the observed number of adult current smokers (top) and the observed BRFSS sample size (bottom) in Washington State zip codes in 2006. County boundaries are indicated.

Fig. 3.
For 2006 Washington BRFSS data: effective sample sizes versus observed sample sizes.

Fig. 4. Smoking prevalence estimates across Washington State zip codes in 2006, using various approaches.

Fig. 5. Comparison of estimated smoking prevalence (left) and estimated smoking counts (right) across zip codes under the spatial logit and unadjusted binomial models.

Fig. 6.
Predicted total adult smokers by zip code in Washington State in 2006, under the spatial logit normal model. County boundaries are indicated.

Fig. 7.
The 95% interval of the predicted total adult smokers by zip code in Washington State in 2006, under the spatial logit normal model.

Table 1

Summary statistics for population data, and the 2006 Washington State BRFSS data on adult current smokers, across zip codes.

	Mean	S.D.	Median	Min	Max
Population	12 570.0	12 931.0	7208.0	11.0	55 700.0
Sample sizes	46.9	54.8	30.0	1.0	384.0
Number of current adult smokers	7.5	9.5	4.0	0.0	67.0

Table 2

Diabetes prevalence rates p_{ij} in area i , $i = 1, \dots, I$, and by post-stratification group, $j = 1, \dots, 6$, corresponding to age and gender. In scenarios 1, 2, 3 and 5 the rates are fixed across areas and the values listed are based on the National Surveillance Data from the CDC (Chen et al., submitted for publication). In scenario 4 the values vary, with spatial structure, across areas, with the first figure in each cell denoting the median rate, and the figures in parentheses a 95% range.

		Age		
		18–44	45–74	75+
Female	1, 2, 3, 5	0.017	0.15	0.17
	4	0.017 (0, 0.034)	0.15 (0.085, 0.21)	0.17 (0, 0.32)
Male	1, 2, 3, 5	0.014	0.16	0.19
	4	0.014 (0, 0.027)	0.16 (0.089, 0.23)	0.19 (0, 0.33)

Table 3

Response rates q_{ij} in area i , $i = 1, \dots, I$ and by age and gender groups, $j = 1, \dots, 6$. In scenarios 1 and 4 there is full response. In scenario 2 the response rates are fixed across areas but vary by group. In scenario 3 the response rates vary, without spatial structure, across areas, with the first figure denoting the median rate, and the figures in parentheses a 95% range. In scenario 5 the response rates vary, with spatial structure, across areas, with the first figure in each cell denoting the median rate, and the figures in parentheses a 95% range.

		Scenario	Age		
			18–44	45–74	75+
Female	1, 4		1	1	1
	2		0.55	0.65	0.8
	3		0.55 (0.38, 0.70)	0.65 (0.48, 0.79)	0.80 (0.67, 0.89)
	5		0.55 (0.46, 0.65)	0.65 (0.57, 0.74)	0.80 (0.74, 0.86)
Male	1, 4		1	1	1
	2		0.50	0.60	0.75
	3		0.50 (0.34, 0.66)	0.60 (0.43, 0.75)	0.75 (0.60, 0.86)
	5		0.50 (0.41, 0.60)	0.60 (0.51, 0.69)	0.75 (0.68, 0.82)

Table 4

Simulation results to examine non-response bias. Tables 2 and 3 give the prevalence and response parameters that change across scenarios. Non-hierarchical unadjusted use the observed y_i and m_i and non-hierarchical adjusted use the Horvitz–Thompson estimator.

(×10 ³)	Non-hierarchical		Unadjust binom		Logit normal		Pseudo-likelihood		Arcsin sqrt		Effect samp size	
	Unadjust	Adjust	Indep	Spatial	Indep	Spatial	Indep	Spatial	Indep	Spatial	Indep	Spatial
<i>Bias²</i>												
Scenario 1	2.2	2.7	26.0	18.9	54.9	40.7	32.1	22.4	9.3	10.8	15.6	12.2
Scenario 2	14.2	2.5	56.0	42.9	54.0	40.0	34.4	24.1	10.0	9.8	17.7	13.4
Scenario 3	15.4	7.1	51.3	38.2	52.5	37.9	32.5	22.3	9.0	10.4	15.7	11.9
Scenario 4	2.2	2.9	36.7	23.6	64.9	42.8	43.7	27.1	18.8	16.1	25.2	17.8
Scenario 5	12.9	2.7	55.2	41.7	53.7	39.1	33.9	23.6	9.9	10.2	17.3	13.2
<i>Variance</i>												
Scenario 1	233.8	220.8	6.5	7.5	3.0	4.8	4.5	6.0	25.8	26.5	15.5	15.4
Scenario 2	253.8	210.3	6.1	7.6	2.8	4.7	2.9	4.8	20.9	23.1	12.2	12.6
Scenario 3	252.7	210.4	8.3	9.1	3.8	5.3	4.4	5.7	25.3	25.4	15.8	15.3
Scenario 4	236.1	222.8	10.0	10.2	5.1	6.8	7.2	8.1	29.6	27.8	19.8	18.6
Scenario 5	249.1	206.1	5.8	7.2	2.6	4.4	2.9	4.6	20.9	22.6	12.0	12.2
<i>MSE</i>												
Scenario 1	236.0	223.4	32.6	26.4	57.9	45.6	36.7	28.5	35.1	37.3	31.1	27.6
Scenario 2	268.0	212.8	62.1	50.5	56.7	44.7	37.4	28.9	30.9	32.9	29.9	26.0
Scenario 3	268.2	217.5	59.6	47.4	56.3	43.2	36.9	28.0	34.3	35.7	31.4	27.2
Scenario 4	238.3	225.7	46.7	33.8	70.1	49.6	50.8	35.2	48.3	43.9	45.0	36.3
Scenario 5	262.0	208.8	61.0	48.9	56.3	43.6	36.9	28.2	30.8	32.8	29.3	25.4

Table 5

Simulation results to examine the effect of selection bias. Non-hierarchical unadjusted use the observed y_i and m_i and non-hierarchical adjusted use the Horvitz–Thompson estimator. Selection is based on $s = \Pr(Z_{ik} = 1 | Y_{ik} = 1)$ where Z_{ik} is a binary variable upon which sampling is based.

(×10 ³)	Non-hierarchical		Unadjusted binomial		Logit normal		Pseudo-likelihood		Arcsin sqrt		Effect samp size	
	Unadjusted	Adjusted	Indep	Spatial	Indep	Spatial	Indep	Spatial	Indep	Spatial	Indep	Spatial
<i>Bias²</i>												
$s = 0.1$	4.4	6.3	23.5	17.6	5.8	7.7	19.4	14.9	20.1	26.5	5.8	5.5
$s = 0.3$	591.7	5.0	806.9	712.2	9.7	11.4	22.7	16.2	27.8	39.1	8.7	11.1
$s = 0.5$	1726.0	3.0	2141.6	1964.2	10.7	13.6	31.3	21.7	27.2	41.8	11.8	17.3
$s = 0.8$	3490.5	1.7	4116.1	3858.8	27.8	19.0	44.9	30.2	17.6	22.6	18.6	19.6
<i>Variance</i>												
$s = 0.1$	315.7	511.2	8.2	8.8	214.8	128.7	100.0	96.5	167.6	161.0	214.8	210.1
$s = 0.3$	473.0	425.5	15.0	15.4	116.6	110.9	45.9	43.0	122.3	112.2	147.3	139.0
$s = 0.5$	532.8	322.4	8.5	12.0	51.8	46.3	8.6	9.3	58.3	49.7	65.1	57.2
$s = 0.8$	544.5	167.8	1.8	7.5	0.6	2.3	0.4	1.4	4.3	6.6	2.3	3.2
<i>MSE</i>												
$s = 0.1$	320.1	517.4	31.7	26.5	220.6	136.4	119.4	111.4	187.7	187.6	220.6	215.7
$s = 0.3$	1064.7	430.6	821.9	727.6	126.3	122.3	68.5	59.2	150.2	151.3	156.0	150.1
$s = 0.5$	2258.8	325.4	2150.2	1976.2	62.5	59.9	39.9	30.9	85.5	91.5	76.9	74.5
$s = 0.8$	4035.0	169.5	4117.9	3866.3	28.4	21.3	45.3	31.6	21.9	29.1	20.9	22.8

Table 6

Model validation on the total for zip code 98801. The true total is estimated to be 5085. The estimates are based on samples of size n .

$(\times 10^3)$	Non-hierarchical		Unadjust binomial		Logit normal		Pseudo-likelihood		Arcsin sqrt		Effect samp size	
	Unadjust	Adjust	Indep	Spatial	Indep	Spatial	Indep	Spatial	Indep	Spatial	Indep	Spatial
<i>Bias²</i>												
$n = 10$	1459.4	126.6	392.3	763.0	11.0	470.5	4.1	268.4	671.5	1671.7	386.1	1178.3
$n = 30$	1539.4	33.7	560.9	839.6	26.2	446.9	10.2	171.6	582.1	1405.0	344.1	1000.0
$n = 50$	1433.5	3.8	657.1	878.0	12.2	355.1	2.4	90.2	273.9	867.8	162.8	643.2
<i>Variance</i>												
$n = 10$	9419.8	21 804.1	193.1	53.2	391.4	143.9	1479.7	772.1	1007.2	549.6	997.7	508.7
$n = 30$	2713.4	7982.3	287.4	100.9	619.5	258.0	2180.4	1447.6	1559.1	983.4	1358.8	799.8
$n = 50$	1600.4	4511.8	312.4	128.5	619.2	282.6	1864.8	1374.0	1193.5	786.8	1124.4	711.9
<i>MSE</i>												
$n = 10$	10 879.3	21 930.6	585.4	816.2	402.4	614.4	1483.8	1040.5	1678.8	2221.4	1383.8	1687.0
$n = 30$	4252.8	8016.0	848.3	940.5	645.7	704.9	2190.6	1619.3	2141.2	2388.4	1702.8	1799.7
$n = 50$	3033.8	4515.7	969.5	1006.5	631.5	637.7	1867.2	1464.1	1467.4	1654.6	1287.3	1355.1

Table 7

Model validation on the total for zip code 98802. The true total is estimated to be 2612. The estimates are based on samples of size n .

($\times 10^3$)	Non-hierarchical		Unadjust binomial		Logit normal		Pseudo-likelihood		Arcsin sqrt		Effect samp size	
	Unadjust	Adjust	Indep	Spatial	Indep	Spatial	Indep	Spatial	Indep	Spatial	Indep	Spatial
<i>Bias²</i>												
$n = 10$	0.8	0.0	110.3	125.8	371.1	289.3	291.7	277.7	14.4	0.1	56.5	18.4
$n = 30$	0.0	1.8	73.1	95.5	246.1	218.5	133.2	148.1	3.1	0.9	29.0	9.2
$n = 50$	0.1	0.0	48.2	72.2	164.6	162.5	67.8	84.4	3.7	0.0	20.8	7.7
<i>Variance</i>												
$n = 10$	3938.0	5608.7	80.4	29.0	115.8	50.9	330.7	194.2	358.1	239.4	304.4	187.3
$n = 30$	1374.3	2165.4	148.3	65.0	203.5	102.3	550.7	388.7	586.4	445.7	475.8	337.6
$n = 50$	729.0	1181.1	142.9	71.1	205.7	113.2	453.0	346.4	461.0	364.7	389.3	292.6
<i>MSE</i>												
$n = 10$	3938.8	5608.7	190.7	154.8	486.9	340.2	622.4	471.9	372.5	239.5	360.9	205.7
$n = 30$	1374.4	2167.2	221.4	160.5	449.6	320.7	683.9	536.8	589.6	446.5	504.8	346.8
$n = 50$	729.1	1181.1	191.1	143.3	370.2	275.6	520.8	430.8	464.8	364.7	410.1	300.3

Table 8

Model validation on the total for zip code 99347. The true total is estimated to be 360. The estimates are based on samples of size n .

(×10 ³)	Non-hierarchical		Unadjust binomial		Logit normal		Pseudo-likelihood		Arcsin sqrt		Effect samp size	
	Unadjust	Adjust	Indep	Spatial	Indep	Spatial	Indep	Spatial	Indep	Spatial	Indep	Spatial
<i>Bias</i> ²												
$n = 10$	3.1	0.9	4.0	3.0	1.1	1.4	0.9	1.2	5.5	6.9	3.9	4.8
$n = 30$	3.6	0.3	3.9	3.0	1.2	1.5	0.6	0.7	4.4	5.7	3.3	4.1
$n = 50$	3.4	0.1	3.7	3.0	1.2	1.5	0.3	0.5	3.3	4.4	2.5	3.2
<i>Variance</i>												
$n = 10$	44.6	84.7	0.9	0.3	1.6	0.6	6.2	3.5	4.2	2.7	4.3	2.6
$n = 30$	13.8	38.1	1.6	0.6	2.4	1.1	11.3	8.0	5.2	3.8	4.9	3.4
$n = 50$	8.0	25.4	1.7	0.7	2.7	1.3	11.2	8.6	4.6	3.4	4.6	3.3
<i>MSE</i>												
$n = 10$	47.7	85.5	5.0	3.2	2.7	2.1	7.1	4.7	9.7	9.6	8.2	7.3
$n = 30$	17.3	38.4	5.4	3.6	3.7	2.6	11.9	8.7	9.7	9.4	8.2	7.4
$n = 50$	11.4	25.6	5.4	3.7	3.9	2.8	11.5	9.0	7.9	7.8	7.1	6.5

Table 9

Model validation across the three zip codes 98801, 98802 and 99347. The estimates are based on samples of size n .

(×10 ³)	Non-hierarchical		Unadjust binomial		Logit normal		Pseudo-likelihood		Arcsin sqrt		Effect samp size	
	Unadjust	Adjust	Indep	Spatial	Indep	Spatial	Indep	Spatial	Indep	Spatial	Indep	Spatial
<i>Bias²</i>												
$n = 10$	1463.3	127.4	506.6	891.7	383.2	761.3	296.8	547.3	691.5	1678.7	446.5	1201.5
$n = 30$	1543.0	35.9	637.8	938.1	273.5	666.8	143.9	320.5	589.6	1411.5	376.3	1013.2
$n = 50$	1436.9	4.0	709.1	953.2	178.0	519.0	70.6	175.1	280.9	872.3	186.2	654.1
<i>Variance</i>												
$n = 10$	13 402.4	27 497.5	274.5	82.5	508.8	195.4	1816.6	969.9	1369.5	791.7	1306.3	698.5
$n = 30$	4101.5	10 185.7	437.3	166.5	825.4	361.4	2742.5	1844.3	2150.8	1432.9	1839.6	1140.7
$n = 50$	2337.4	5718.3	457.0	200.3	827.6	397.1	2328.9	1728.9	1659.1	1154.8	1518.3	1007.8
<i>MSE</i>												
$n = 10$	14 865.7	27 624.9	781.1	974.2	892.0	956.7	2113.4	1517.2	2061.0	2470.4	1752.8	1900.1
$n = 30$	5644.5	10 221.6	1075.1	1104.6	1098.9	1028.2	2886.4	2164.8	2740.4	2844.4	2215.9	2154.0
$n = 50$	3774.3	5722.3	1166.0	1153.5	1005.6	916.1	2399.5	1904.0	1940.0	2027.1	1704.5	1661.9

Table 10

Comparison of posterior medians of spatial and non-spatial standard deviations, σ_u and σ_v , and proportion of total variance that is spatial p_s , for three different priors on the spatial precision σ_u^{-2} for each of the hierarchical models described in Section 2.

	Prior for σ_u^{-2}	Unadjust binomial	Logit normal	Pseudo-likelihood	Arcsin sqrt	Effect samp size
σ_u	Ga(0.50, 0.008)	0.35	0.30	0.31	0.065	0.29
	Ga(1.00, 0.026)	0.34	0.29	0.30	0.079	0.28
	Ga(0.35, 0.001)	0.35	0.31	0.31	0.053	0.29
σ_v	Ga(0.50, 0.008)	0.17	0.25	0.38	0.078	0.41
	Ga(1.00, 0.026)	0.17	0.25	0.38	0.073	0.41
	Ga(0.35, 0.001)	0.16	0.25	0.38	0.081	0.41
p_s	Ga(0.50, 0.008)	0.79	0.76	0.78	0.12	0.77
	Ga(1.00, 0.026)	0.78	0.75	0.76	0.14	0.75
	Ga(0.35, 0.001)	0.79	0.77	0.78	0.099	0.78