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

Article

Hierarchical and empirical Bayes small domain estimation of the proportion of persons without health insurance for minority subpopulations

by Malay Ghosh, Dalho Kim, Karabi Sinha, Tapabrata Maiti, Myron Katzoff and Van L. Parsons

June 2009

Statistics Canada Statistique Canada

Canada

Survey Methodology, June 2009
Vol. 35, No. 1, pp. 53-66
Statistics Canada, Catalogue No. 12-001-X

Hierarchical and empirical Bayes small domain estimation of the proportion of persons without health insurance for minority subpopulations

Malay Ghosh, Dalho Kim, Karabi Sinha, Tapabrata Maiti, Myron Katzoff and Van L. Parsons [1]

Abstract

The paper considers small domain estimation of the proportion of persons without health insurance for different minority groups. The small domains are cross-classified by age, sex and other demographic characteristics. Both hierarchical and empirical Bayes estimation methods are used. Also, second order accurate approximations of the mean squared errors of the empirical Bayes estimators and bias-corrected estimators of these mean squared errors are provided. The general methodology is illustrated with estimates of the proportion of uninsured persons for several cross-sections of the Asian subpopulation.

Key Words: Asian; Bias-corrected; Mean squared error; Second order accurate.

1. Introduction

The main motivation behind this work was small domain estimation of the proportion of individuals without health insurance for different minority subpopulations. The small domains were constructed on the basis of age, sex, race and the region where the person belongs. The National Health Interview Survey (NHIS) data provide the individual level binary response (that is whether or not a person has health insurance) along with individual level covariates. The data can be obtained at http://www.cdc.gov/nchs/nhis.htm. The design of NHIS is discussed in Botman, Moore, Moriarity and Parsons (2000).

In a typical year the NHIS samples dwelling units, the collective members of each unit being referred to as a household, and members with a "strong" relationship being referred to as a family. (Structural units are more explicitly defined in Chapter 5.2 in the Census document at www.census.gov/prod/2002pubs/tp63rv.pdf). Each year the NHIS data contain about 40,000 households, of which over 98% are one-family households, and contain about 100,000 persons. For "family-type" questions, *e.g.*, on insurance coverage, all adults at home are invited to participate in the interview, but proxy adult response is also allowed. Children require an adult proxy.

The original survey for any given year contains data on more than 100,000 individuals and on over 800 variables. Of these individuals, we have information on the primary response variable, namely whether a person has health insurance or not. In addition, there is information on demographic characteristics such as age, sex, race, region, education, income status, medical condition, disability conditions (if any) and many other socio-economic factors.

For the entire US population, the direct estimates for these domains, namely the weighted sample proportions, are fairly reliable, since the sample size for each domain is reasonably large. This need not be the case though when our analysis is targeted towards specific subpopulations, such as Hispanics, Asians and similar minority sectors of the community.

For a targeted minority subpopulation, the sample size in a domain is not always very large. Hence, the direct estimates may not be very reliable, being accompanied with large standard errors and coefficients of variation. This calls for the use of small domain estimation techniques, where indirect estimates are obtained for these domains based on implicit or explicit models. These models help building a link between these domains, and thus produce typically estimates of greater precision by borrowing strength.

We employ both hierarchical Bayes (HB) and empirical Bayes (EB) methodology to obtain small domain estimates and find also the associated measures of precision. The analysis is based on a HB analogue of the generalized linear mixed model (GLMM) to obtain posterior means and posterior standard errors of the population small domain proportions. The method was proposed in Ghosh, Natarajan, Stroud and Carlin (1998). The EB approach is based on the theory of optimal estimating functions. We obtain EB estimators and the corresponding approximate mean squared error estimators by an asymptotic method analogous to that of Prasad and Rao (1990) and Ghosh and Maiti (2004). While the procedure of Ghosh and Maiti (2004) is based on area-level data, the present approach uses unit level data. Hence, by necessity, one needs some modification of the procedure proposed in Ghosh and Maiti (2004) in developing the estimators. Also, the general methodology,

1. Malay Ghosh, University of Florida; Dalho Kim, Kyungpook National University; Karabi Sinha, University of California at Los Angeles; Tapabrata Maiti, Michigan State University; Myron Katzoff, National Center for Health Statistics; Van L. Parsons, National Center for Health Statistics.

like that of Ghosh and Maiti is not restricted only to binary data. The methodology is applicable to the natural exponential family with quadratic variance functions. (Morris 1982, 1983). The development of mean squared errors of the estimates under the proposed model is somewhat simpler than that of Ghosh and Maiti (2004) for the binary case. Moreover, like Ghosh and Maiti (2004), our analysis utilizes the survey weights along with the model to derive the small domain estimates. Thus, our method, in some sense, can be regarded as design-assisted model-based estimation.

Survey weights attached to individual sampling units are usually proportional to inverses of their selection probabilities. They are often used to produce design-unbiased estimators. The classic example is the celebrated Horvitz-Thompson estimator. However, while such estimators guard against model failure, they may result in loss of efficiency if the assumed model is true. For example, in a simple Bayesian set up, if $y_i \mid \theta_i$ are independently distributed $\mathrm{N}(\theta_i, 1)$, while θ_i are independently and identically distributed $\mathrm{N}(\mu, A), (i = 1, \ldots, n)$, then the Bayes estimator (posterior mean) of $\bar{\theta} = n^{-1}\sum_{i=1}^{n}\theta_i$ is $n^{-1}\sum_{i=1}^{n}[(1-B)y_i + B\mu] = (1-B)\bar{y} + B\mu$, where $B = (1+A)^{-1}$. This estimator has Bayes risk $n^{-1}(1-B)$ under the assumed model. On the other hand, the estimator $\sum_{i=1}^{n} w_i y_i$ of $\bar{\theta}$, with $\sum_{i=1}^{n} w_i = 1$ has Bayes risk $n^{-1}(1-B) + E[(1-B)\bar{y} + B\mu - \sum_{i=1}^{n} w_i y_i]^2$. If, however, the assumed model is not true, for example, θ_i are independently and identically distributed $\mathrm{N}(\mu, A), (i = 1, \ldots, n)$, where A departs widely from A_0, then the Bayes risk of the estimator $(1-B)\bar{y} + B\mu$ of $\bar{\theta}$ has Bayes risk $n^{-1}(1-B_0) + (B-B_0)^2(\bar{y}-\mu)^2, B_0 = (1+A_0)^{-1}$, which can be quite larger than the corresponding Bayes risk of $\sum_{i=1}^{n} w_i y_i$ depending of course on B_0, μ and the w_i.

The present paper produces small domain estimates of the proportion of uninsured persons for the Asian population. The estimates and measures of precision are based both on the hierarchical Bayesian model as well as the EB model. The analysis was done for all the individual years 1997-2000. For brevity, the results are reported only for the year 2000. We carried out a similar analysis for the Hispanic population also. In this case, the number of small domains was 336. Since the methodology was the same as that for the Asians, to save space, we have not included in this paper that analysis as well.

The Asian group is formally composed of the (1) Chinese, (2) Filipino, (3) Asian Indian, and (4) others such as Koreans, Vietnamese, Japanese, Hawaiian, Samoan, Guamanian *etc*. These individuals are assigned to specific domains depending on their age, race, gender and the region they come from. There are 3 age-groups (0-17, 18-64 and 65+), 2 Genders, 4 Races and 4 Regions depending on the size of the Metropolitan Statistical Area ($<$499,999; 500,000-999,999; 1,000,000-2,499,999 $>$2,500,000). Thus, the total number of domains equals $3 \times 2 \times 4 \times 4 = 96$. When the individuals are distributed to their respective domains, it turns out that many of the domains contain only a few observations. Indeed, there are several domains with a sample of size 1, while one domain has sample size zero.

The outline of the remaining sections is as follows. Section 2 addresses the selection of covariates for the Asians. Section 3 discusses the general HB methodology needed for obtaining the small domain estimates and the associated measures of precision. Section 4 discusses the adequacy of the proposed HB model. Section 5 discusses an alternative EB methodology, finds second order correct (to be made precise later) mean squared errors (MSE's) of the proposed EB estimators, and also second order correct approximation of these MSE's. Section 6 finds the small domain estimates and the corresponding measures of precision for the Asian subpopulation in 2000 using both the HB and the EB methodology, and these estimates are compared with the direct estimates. Some concluding remarks are made in Section 7.

2. Selection of covariates

As mentioned in the introduction, the number of covariates exceeds 800. Inclusion of all of them in the initial model is impractical and unnecessary. We started with what we deemed to be a meaningful set of 6 covariates and used a fully stepwise selection process (with a significance level of 0.05) to finally come up with the best model.

The six covariates that we considered were: (1) legal marital status, (2) family size, (3) education level, (4) total earnings from the previous year, (5) total family income, and (6) full time working status.

After the stepwise procedure, our final model included, along with the intercept term, the covariates family size, education level, and total family income.

Since the SURVEYREG procedure in SAS Version 8 fits linear regression models and produces hypothesis tests and estimates for survey data, we used this procedure for our covariate selection. Logistic regression for covariate selection was not available at the time when this research was done. It may be noted though that SURVEYREG acounts for clustering and unequal weighting, and produces standard errors that correctly account for complex survey designs.

3. Hierarchical Bayesian analysis

A general one-parameter exponential family model is given by

$$f(y_{ij} \mid \theta_{ij}) = \exp[\xi_{ij}\{y_{ij}\theta_{ij} - \psi(\theta_{ij})\}]h(y_{ij}; \xi_{ij}), \qquad (3.1)$$

$j = 1, \ldots, n_i, i = 1, \ldots, k$. Here y_{ij} is the response of the j^{th} unit in the i^{th} small domain, while ξ_{ij}, the "so-called" overdispersion parameters are assumed to be known, and are taken as 1 without loss of generality. This is because one can otherwise work with the transformed parameters. $\zeta_{ij} = \xi_{ij}\theta_{ij}$. The function h is a positive function which depends on the y_{ij}, but not on the θ_{ij}. If y_{ij} is binary with success probability p_{ij}, then $\theta_{ij} = \text{logit}(p_{ij})$. In our example, y_{ij}, the response of the j^{th} individual in the i^{th} small domain, is 1 or 0 depending on whether the person does not or does have health insurance. We are interested in estimation of $\bar{\mu}_{iw} = \sum_{j=1}^{n_i} w_{ij} p_{ij}$, the domain specific weighted averages of the population proportions. In this case, the direct estimator of $\bar{\mu}_{iw}$ is $\sum_{j=1}^{n_i} w_{ij} y_{ij}$. These direct estimators are usually subject to large standard errors and coefficients of variation. The survey weights w_{ij} are assumed to be known, and are normalized so that $\sum_{j=1}^{n_i} w_{ij} = 1$ for all $i = 1, \ldots, k$. It must be admitted though that often in practice, the w_{ij} are only estimates, for example taking into account post-stratification and non-response. However, the actual mechanism used to generate these weights are unavailable to secondary users of the data, and we need to assume the weights to be known. Another important example, not specifically considered in this paper, is $y_{ij} \sim \text{Poisson}(\lambda_{ij})$, so that $\theta_{ij} = \log(\lambda_{ij})$. One can use a Poisson model here based on the domain level counts of uninsured people The difficulty lies in the fact that in the present example, we have individual level and *not* domain level covariates. Modelling the counts via domain-level covariates is not possible in this situation.

In this section, we discuss how to carry out the analysis for the general hierarchical Bayesian model when we are interested in estimating $\mu_{ij} = E(y_{ij} \mid \theta_{ij}) = \psi'(\theta_{ij})$. Since $\psi''(\theta_{ij}) = \text{var}(y_{ij} \mid \theta_{ij})$, μ_{ij} is a one-to-one function of θ_{ij}. In particular, $\mu_{ij} = p_{ij}$ in the binary case. Specific applications will be considered in Section 5.

The next stage of the model is

$$\theta_{ij} = \boldsymbol{x}_{ij}^T \boldsymbol{b} + u_i;\ j = 1, \ldots, n_i, i = 1, \ldots, k, \qquad (3.2)$$

where $\boldsymbol{x}_{ij}$ are the design vectors, or equivalently the predictor vectors, $\boldsymbol{b}$ is the vector of regression parameters, and u_i are the random effects. It is assumed that u_i are iid $\text{N}(0, \sigma_u^2)$. Also, let $\boldsymbol{X}^T = (\boldsymbol{x}_{11}, \ldots, \boldsymbol{x}_{1n_1}, \ldots, \boldsymbol{x}_{k1}, \ldots, x_{kn_k})$, and assume that $\boldsymbol{X}$ is a full rank matrix.

Finally, it is assumed that $\boldsymbol{b}$ and σ_u^2 are mutually independent, where $\boldsymbol{b}$ has the improper uniform prior on, R^P, and σ_u^2 has an inverse gamma distribution with parameters $c/2.d/2.i.e.$, $\pi(\sigma_u^2) \propto \exp(-c/2\sigma_u^2)(\sigma_u^2)^{-d/2-1}$, $c > 0$.

Let $\boldsymbol{y} = (y_{11}, \ldots, y_{1n_1}, \ldots, y_{k1}, \ldots, y_{kn_k})^T$, and $\boldsymbol{\theta} = (\theta_{11}, \ldots, \theta_{1n_1}, \ldots, \theta_{k1}, \ldots, \theta_{kn_k})^T$. Then the joint posterior is given by

$$\pi(\boldsymbol{\theta}, \boldsymbol{b}, \sigma_u^2 \mid \boldsymbol{y}) \propto \prod_{i=1}^{k}\prod_{j=1}^{n_i} f(y_{ij} \mid \theta_{ij})$$
$$\times (\sigma_u^2)^{-k/2} \exp\left[-\frac{1}{2\sigma_u^2}\sum_{i=1}^{m}\sum_{j=1}^{n_i}(\theta_{ij} - \boldsymbol{x}_{ij}^T\boldsymbol{b})^2\right]$$
$$\times (\sigma_u^2)^{-d/2-1} \exp\left(-\frac{c}{2\sigma_u^2}\right). \qquad (3.3)$$

This is a nonconjugate Bayesian analysis, and is not implementable analytically. Instead, we use the Markov chain Monte Carlo (MCMC) numerical integration technique. In particular, we employ the Gibbs sampler. The general MCMC technique is discussed in many places. A convenient reference is Tanner (1996, Chapter 6).

In order to implement the Gibbs sampler, we need to find the full conditionals of $\theta_{ij}, \boldsymbol{b}$ and σ_u^2. The full conditionals are given by

$$\sigma_u^2 \mid \boldsymbol{\theta}, \boldsymbol{b}, \boldsymbol{y} \sim \text{IG}\left(\frac{\sum_{i=1}^{k}\sum_{j=1}^{n_i}(\theta_{ij} - \boldsymbol{x}_{ij}^T\boldsymbol{b})^2 + c}{2}, \frac{k+d}{2}\right);$$

$$\boldsymbol{b} \mid \boldsymbol{\theta}, \sigma_u^2, \boldsymbol{y} \sim N((\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{y}, \sigma_u^2(\boldsymbol{X}^T\boldsymbol{X})^{-1});$$

$$\theta_{ij} \mid \boldsymbol{b}, \sigma_u^2, \boldsymbol{y} \sim f(y_{ij} \mid \theta_{ij}) \exp\left[-\frac{1}{2\sigma_u^2}(\theta_{ij} - \boldsymbol{x}_{ij}^T\boldsymbol{b})^2\right].$$

Our data analysis is based on generating samples from the above conditionals specialized to the binary case. Generation of samples from the conditionals of σ_u^2 and $\boldsymbol{b}$ is standard. This is not so for the θ_{ij}, and requires the Metropolis-Hastings algorithm. For a discussion of this algorithm, we refer once again to Tanner (1996).

If $\hat{\mu}_{ij}^{(r)}$ denotes the sampled value of μ_{ij} generated from the r^{th} draw, and the number of draws is R, then the Monte Carlo estimate of $E(\mu_{ij} \mid \boldsymbol{y})$ is $R^{-1}\sum_{r=1}^{R}\hat{\mu}_{ij}^{(r)}$. Similarly, the Monte-Carlo estimate of $\text{var}(\mu_{ij} \mid \boldsymbol{y})$ is $R^{-1}\sum_{r=1}^{R}(\hat{\mu}_{ij}^{(r)})^2 - (R^{-1}\sum_{r=1}^{R}\hat{\mu}_{ij}^{(r)})^2$. Finally, Monte-Carlo estimate of $\text{cov}(\mu_{ij}, \mu_{i'j'} \mid \boldsymbol{y})$ is given by $R^{-1}\sum_{r=1}^{R}(\hat{\mu}_{ij}^{(r)}\hat{\mu}_{i'j'}^{(r)}) - (R^{-1}\sum_{r=1}^{R}\hat{\mu}_{ij}^{(r)})$ $(R^{-1}\sum_{r=1}^{R}\hat{\mu}_{i'j'}^{(r)})$. Based on these calculations, it is now immediate to find $E[\bar{\mu}_{iw} \mid \boldsymbol{y}] = \sum_{j=1}^{n_i} w_{ij} E(\mu_{ij} \mid \boldsymbol{y})$ and $V[\bar{\mu}_{iw} \mid \boldsymbol{y}] = \sum_{j=1}^{n_i} w_{ij}^2 V(\mu_{ij} \mid \boldsymbol{y}) + \sum_{1 \le j \ne j' \le n_i} w_{ij} w_{ij'}$ $\text{Cov}(\mu_{ij}, \mu_{ij'} \mid \boldsymbol{y})$. In contrast, the direct unbiased estimator of $\bar{\mu}_{iw}$ is given by $\bar{y}_{iw} = \sum_{j=1}^{n_i} w_{ij} y_{ij}$. However, as noted earlier, for many of these domains, the sample sizes are so small that these unbiased estimators are subject to large standard errors and coefficients of variation.

4. Empirical Bayes estimation

Once again, let y_{ij} denote the response of the j^{th} unit in the i^{th} small domain $(j = 1, \ldots, n_i; i = 1, \ldots, k)$. Also, we assume the exponential family model for the y_{ij} as given in (3.1), but it is assumed in addition that the y_{ij} has a probability function or a probability density function belonging to the natural exponential family quadratic variance function (NEF-QVF) class. We may recall that $\mu_{ij} = E(y_{ij} \mid \theta_{ij}) = \psi'(\theta_{ij})$. With the quadratic variance function structure, $\text{Var}(y_{ij} \mid \theta_{ij}) = Q(\mu_{ij}) = v_0 + v_1\mu_{ij} + v_2\mu_{ij}^2$, where v_0, v_1 and v_2 are not simultaneously zero. Morris (1982, 1983) has characterized distributions belonging to the NEF-QVF family. The family consists of the six basic distributions, namely, (i) Bernoulli, (ii) Poisson, (iii) normal with known variance, (iv) geometric, (v) exponential, (vi) hyperbolic secant, and their convolutions. In this way, binomial, negative binomial and gamma distributions also belong to this family. For the Bernoulli distribution, $v_0 = 0, v_1 = 1$ and $v_2 = -1$. For the Poisson distribution, $v_0 = v_2 = 0$ and $v_1 = 1$. For the normal distribution with known variance $\sigma^2, \xi_{ij} = \sigma^{-2}, v_0 = 1$ and $v_1 = v_2 = 0$. Once again we will assume without loss of generality that $\xi_{ij} = 1$.

We propose in this section EB estimators of the small domain means. To this end, we begin with the general NEF-QVF family of distributions along with a conjugate prior for the canonical parameter of the exponential model. Together they constitute an overdispersed NEF-QVF family of distributions. Specifically, we consider the conjugate prior with pdf

$$\pi(\theta_{ij}) = \exp[\lambda\{m_{ij}\theta_{ij} - \psi(\theta_{ij})\}]C(\lambda, m_{ij}) \qquad (4.1)$$

for θ_{ij}, where $m_{ij} = g(\boldsymbol{x}_{ij}^T \boldsymbol{b}), j = 1, \ldots, n_i; i = 1, \ldots, k$. Here $\boldsymbol{x}_{ij}$ is the design vector associated with the j^{th} unit in the i^{th} small domain, and g is the link function. Then (Morris 1983),

$$E(\mu_{ij}) = m_{ij}; \text{var}(\mu_{ij}) = Q(m_{ij})/(\lambda - v_2), \qquad (4.2)$$

where we assume that $\lambda > \max(0, v_2)$. Since $\text{var}(\mu_{ij})$ is strictly decreasing in λ, we may interpret the latter as the precision parameter.

We first obtain the Bayes estimator of μ_{ij}. This is given by (Morris 1983)

$$E(\mu_{ij} \mid y_{ij}) = \frac{1}{\lambda + 1} y_{ij} + \frac{\lambda}{\lambda + 1} m_{ij}(\boldsymbol{b}).$$

The above can also be viewed as the best linear unbiased predictor (BLUP) of μ_{ij}. To see this, we calculate

$$E(y_{ij}) = E(\mu_{ij}) = m_{ij}; \text{cov}(y_{ij}, \mu_{ij}) = \text{var}(\mu_{ij})$$
$$= Q(m_{ij})/(\lambda - v_2); \text{var}(y_{ij}) = \frac{\lambda + 1}{\lambda - v_2} Q(m_{ij}).$$

Hence, the BLUP of μ_{ij} is given by

$$m_{ij}(\boldsymbol{b}) + \frac{\text{cov}(y_{ij}, \mu_{ij})}{\text{var}(y_{ij})}(y_{ij} - m_{ij}(\boldsymbol{b}))$$
$$= \frac{1}{\lambda + 1} y_{ij} + \frac{\lambda}{\lambda + 1} m_{ij}(\boldsymbol{b}). \qquad (4.3)$$

Thus the Bayes estimator of $\bar{\mu}_{iw} = \sum_{j=1}^{n_i} w_{ij}\mu_{ij}$ is given by $\sum_{j=1}^{n_i} w_{ij} E(\mu_{ij} \mid y_{ij})$.

In practice, however, $\boldsymbol{b}$ and λ are unknown, and need to be estimated from the marginals of the y_{ij}. However, except for the normal distribution, these marginals are fairly complicated, and finding MLE's from the marginal likelihoods can become quite formidable. Instead, we find estimates based on some optimal unbiased estimating equations (Godambe and Thompson 1989) which requires only evaluation of the first four moments of these marginals. To this end, we begin with the the elementary unbiased estimating functions $ig_{1ij} = y_{ij} - m_{ij}$ and $g_{2ij} = (y_{ij} - m_{ij})^2 - (\lambda + 1)/(\lambda - v_2)V(m_{ij})$. In order to construct the optimal estimating equations, let

$$\boldsymbol{D}_{ij}^T = \begin{bmatrix} -E\left(\dfrac{\partial g_{1ij}}{\partial \boldsymbol{b}}\right) & -E\left(\dfrac{\partial g_{2ij}}{\partial \boldsymbol{b}}\right) \\ -E\left(\dfrac{\partial g_{1ij}}{\partial \lambda}\right) & -E\left(\dfrac{\partial g_{2ij}}{\partial \lambda}\right) \end{bmatrix}.$$

Also, let

$$\boldsymbol{\Sigma}_{ij} = \begin{bmatrix} \mu_{2ij} & \mu_{3ij} \\ \mu_{3ij} & \mu_{4ij} - \mu_{2ij}^2 \end{bmatrix},$$

where $\mu_{rij} = E(y_{ij} - m_{ij})^r$ is the r^{th} central moment of y_{ij} based on its marginal distribution. The optimal estimating equations are then given by $\sum_{i=1}^{k}\sum_{j=1}^{n_i} \boldsymbol{D}_{ij}^T \boldsymbol{\Sigma}_{ij}^{-1} \boldsymbol{g}_{ij} = \boldsymbol{0}$, where $\boldsymbol{g}_{ij} = (g_{1ij}\ g_{2ij})^T$. We obtain estimates of $\boldsymbol{b}$ and λ (if they exist) by solving these equations. The solutions of these equations are found by the Nelder-Meade algorithm.

Unfortunately, the above method fails for binary data. In this case, $v_2 = -1$ so that $\text{var}(y_{ij})$ does not depend on λ. Indeed, the marginal beta-binary distributions of the y_{ij} are unidentifiable in λ. A simple way to verify this is that if $y \mid p \sim \text{Bin}(1, p)$, and $p \sim \text{Beta}(\lambda m, \lambda(1 - m))$, then $E(y) = E(p) = m$, and a binary distribution is completely characterized by its mean. The problem does not occur for a Binomial(n, p) distribution with $n \geq 2$ since with the same marginal for p, the mgf of the marginal distribution of the binomial y is $E[(p\exp(t) + 1 - p)^n]$ which depends on λ.

For binary $y_{ij}, \partial g_{1ij}/\partial\lambda = \partial g_{2ij}/\partial\lambda = 0$ so that the second element of the vector $\sum_{i=1}^{k}\sum_{j=1}^{n_i} \boldsymbol{D}_{ij}^T \boldsymbol{\Sigma}_{ij}^{-1} \boldsymbol{g}_{ij}$ is zero. Accordingly, the proposed estimating equations approach fails to estimate λ. The basic data, to be considered in our application, is binary, and this necessitates modification of the proposed procedure.

We have thus considered the optimal estimating function (for known λ)

$$\sum_{i=1}^{k}\sum_{j=1}^{n_i} [(y_{ij} - m_{ij})/(\text{var}(y_{ij})] \frac{\partial m_{ij}}{\partial \boldsymbol{b}} = \boldsymbol{0},$$

since $\partial g_{1ij}/\partial \boldsymbol{b} = -\partial m_{ij}/\partial \boldsymbol{b}$. It may be noted also that in this case $\text{var}(y_{ij}) = V(m_{ij}) = m_{ij}(1 - m_{ij})$. Also, with the logistic representation, $m_{ij}(\boldsymbol{b}) = \exp(\boldsymbol{x}_{ij}^T \boldsymbol{b})/[1 + \exp(\boldsymbol{x}_{ij}^T \boldsymbol{b})]$, one gets $\partial m_{ij}/\partial \boldsymbol{b} = -m_{ij}(1 - m_{ij})\boldsymbol{x}_{ij}$. Thus $\boldsymbol{b}$ is estimated from the estimating equations $\sum_{i=1}^{k}\sum_{j=1}^{n_i} \boldsymbol{x}_{ij} y_{ij} = \sum_{i=1}^{k}\sum_{j=1}^{n_i} \boldsymbol{x}_{ij} m_{ij}$. Denoting this estimator by $\hat{\boldsymbol{b}}$, the EB estimator of μ_{ij} is given by

$$\hat{\mu}_{ij}^{\text{EB}} = \frac{1}{\lambda + 1} y_{ij} + \frac{\lambda}{\lambda + 1} m_{ij}(\hat{\boldsymbol{b}}). \tag{4.4}$$

Accordingly, the EB estimator of $\bar{\mu}_{iw}$ is $\hat{\bar{\mu}}_{iw}^{\text{EB}} = \sum_{j=1}^{n_i} w_{ij} \hat{\mu}_{ij}^{\text{EB}}$.

The procedure described above assumes a known λ. One can find estimates for the μ_{ij} for different choices of λ. In this article, we have tried $\lambda = 0.1, 0.5$ and 1, and have compared the estimates with the corresponding HB estimates.

Next, in this section, we find the mean squared errors (MSE) and also the estimated MSE's of $\hat{\bar{\mu}}_{iw}^{\text{EB}}$ assuming known λ. We state two theorems in this section. Some notations are needed before stating these theorems. Let $\boldsymbol{M} = \text{Diag}(m_{11}, \ldots, m_{1n_1}, \ldots, m_{k1}, \ldots, m_{kn_k})$ and $\boldsymbol{\Sigma}(\boldsymbol{b}) = \boldsymbol{X}^T \boldsymbol{M} (\boldsymbol{I} - \boldsymbol{M}) \boldsymbol{X} = \sum_{i=1}^{k}\sum_{j=1}^{n_i} m_{ij}(1 - m_{ij}) \boldsymbol{x}_{ij} \boldsymbol{x}_{ij}^T$. Also, let $n_T = \sum_{i=1}^{k} n_i$. It is assumed that $1 \le n_i \le C$ for every *i*, so that $n_T = O_e(k)$, where O_e denotes the exact order. The two theorems are now given below.

Theorem 1. Assume $\boldsymbol{\Sigma}(\boldsymbol{b}) = O_e(k)$, *i.e.*, each element of $\boldsymbol{\Sigma}(\boldsymbol{b})$ is bounded below by some constant C_1, and is bounded above by some constant C_2, where $0 < C_1 < C_2 < \infty$. Then an approximate expression for $\text{MSE}(\hat{\bar{\mu}}_{iw}^{\text{EB}})$ correct up to $O(k^{-1})$ is given by

$$\begin{aligned}\text{MSE}(\hat{\bar{\mu}}_{iw}^{\text{EB}}) \doteq{} & \frac{\lambda}{(\lambda+1)^2} \sum_{j=1}^{n_i} w_{ij}^2 m_{ij}(b)(1 - m_{ij}(b)) \\ & + \frac{\lambda^2}{(\lambda+1)^2} \left[\sum_{j=1}^{n_i} w_{ij} m_{ij}(b)(1 - m_{ij}(b)) \boldsymbol{x}_{ij}\right]^T \\ & \times \boldsymbol{\Sigma}^{-1}(\boldsymbol{b}) \left[\sum_{j=1}^{n_i} w_{ij} m_{ij}(b)(1 - m_{ij}(b)) \boldsymbol{x}_{ij}\right]. \end{aligned} \tag{4.5}$$

Theorem 2. Assume $\boldsymbol{\Sigma}(\boldsymbol{b}) = O_e(k)$. Then the following approximation to $\text{MSE}(\hat{\bar{\mu}}^{\text{EB}})$ holds correct up to $O(k^{-1})$.

$$\begin{aligned} & \frac{\lambda}{(1+\lambda)^2} \sum_{j=1}^{n_i} \Bigg[m_{ij}(\hat{\boldsymbol{b}})(1 - m_{ij}(\hat{\boldsymbol{b}})) - (1 - 2m_{ij}(\hat{\boldsymbol{b}}))m_{ij}(\hat{\boldsymbol{b}}) \\ & \qquad (1 - m_{ij}(\hat{\boldsymbol{b}})) \frac{1}{2} \boldsymbol{\Sigma}^{-1}(\hat{\boldsymbol{b}}) \begin{pmatrix} tr(\boldsymbol{\Sigma}^{-1}(\hat{\boldsymbol{b}}) \boldsymbol{K}_1(\hat{\boldsymbol{b}})) \\ \cdot \\ \cdot \\ \cdot \\ tr(\boldsymbol{\Sigma}^{-1}(\hat{\boldsymbol{b}}) \boldsymbol{K}_p(\hat{\boldsymbol{b}})) \end{pmatrix} \\ & \qquad + m_{ij}^2(\hat{\boldsymbol{b}})(1 - m_{ij}(\hat{\boldsymbol{b}}))^2 \boldsymbol{x}_{ij}^T \boldsymbol{\Sigma}^{-1}(\hat{\boldsymbol{b}}) \boldsymbol{x}_{ij} \Bigg] \\ & \qquad + \frac{\lambda^2}{(\lambda+1)^2} \left[\sum_{j=1}^{n_i} w_{ij} m_{ij}(\hat{\boldsymbol{b}})(1 - m_{ij}(\hat{\boldsymbol{b}})) \boldsymbol{x}_{ij}\right]^T \\ & \qquad \times \boldsymbol{\Sigma}^{-1}(\hat{\boldsymbol{b}}) \left[\sum_{j=1}^{n_i} w_{ij} m_{ij}(\hat{\boldsymbol{b}})(1 - m_{ij}(\hat{\boldsymbol{b}})) \boldsymbol{x}_{ij}\right]. \end{aligned} \tag{4.6}$$

The proofs of these theorems are deferred to the Appendix. We will apply these results in finding approximate estimates of MSE's of EB estimators in the next section. However, before that, the following point is worth noting.

If one denotes the coefficient of $\lambda/(1+\lambda)^2$ by $B_i(\hat{b})$ and the coefficient of $\lambda^2/(1+\lambda)^2$ by $C_i(\hat{b})$ in Theorem 2, then noting that $B_i(\hat{b}) = O(1)$ and $C_i(\hat{b}) = O(k^{-1})$, for large k, $\text{MSE}(\hat{\bar{\mu}})$ is maximized at $\hat{\lambda} = (B_i(\hat{b}))/(B_i(\hat{b}) - 2C_i(\hat{b}))$ which is typically very close to 1. The resulting prior with $\hat{\lambda}$ replacing λ is the data adaptive approximate least favorable prior. In the example to be considered, this estimated λ turns out to be 1.003 which conforms the above observation.

5. Small domain estimates for Asians

We first describe how the small domains are constructed. Consider the 4-tuple (k_1, k_2, k_3, k_4), where $k_1 = 1, 2, 3$ or 4 according as the person is Chinese, Filipino, Asian Indian or Islanders. Next $k_2 = 1$ or 2 according as the person is a male or a female. Then $k_3 = 1, 2$ or 3 according as the person belongs to the age-group 0-17, 18-64 or 65+. Finally, $k_4 = 1, 2, 3$ or 4 according as the person belongs to a Metropolitan Statistical Area (MSA) of size $\le 499{,}999$, 500,000 - 999,999, 1,000,000 - 2,499,999 or $\ge 2{,}500{,}000$. A small domain is now numbered by the formula $24(k_1 - 1) + 12(k_2 - 1) + 4(k_3 - 1) + k_4$ corresponding to the 4-tuple (k_1, k_2, k_3, k_4). For example, the small domain consisting of Filipino females belonging to the age-group

18-64 and a MSA of size $500,000 - 999,999$ is numbered 42.

The basic data consist of $y_{ij} = 1$ or 0 if the j^{th} individual in the i^{th} small domain does not (does) have health insurance;

$\tilde{w}_{ij} =$ the sampling weight attached to the j^{th} unit in the i^{th} small domain;

$w_{ij} = \tilde{w}_{ij} / \sum_{j=1}^{n_i} \tilde{w}_{ij}$ so that $\sum_{j=1}^{n_i} w_{ij} = 1$ for each i.

$x_{ij1} =$ the family size of the j^{th} unit in the i^{th} small domain;

$x_{ij2} =$ the education level of the j^{th} unit in the i^{th} small domain;

$x_{ij3} =$ total family income of the j^{th} unit in the i^{th} small domain;

Let $p_{ij} = E(y_{ij})$. For the HB analysis, we model

$$\theta_{ij} = \text{logit}(p_{ij}) = b_0 + b_1 x_{ij1} + b_2 x_{ij2} + b_3 x_{ij3} + u_i,$$
$$j = 1, \ldots, n_i, \; i = 1, \ldots, 96.$$

The direct domain estimates are given by $\hat{p}_{iw} = \sum_{j=1}^{n_i} w_{ij} y_{ij}$. The corresponding hierarchical Bayes estimates are given by $\hat{p}_{iw}^{\text{HB}} = \sum_{j=1}^{n_i} w_{ij} E(p_{ij} \mid \boldsymbol{y})$. We use MCMC as described in Section 2 to obtain these estimates. They are referred to in the table as HB. The associated posterior standard errors are referred to as $\text{se}(\text{HB})$. Our hyperprior considers: $c = 0.2, 0.02, 0.002$; $d = 0.2, 0.02, 0.002$. The results are very insensitive to the choice of the hyperpriors, and are reported only for $c = d = 0.02$. In addition, we have EB estimators for different choices of the parameter λ. The results are reported for $\lambda = 0.1, 0.5$ and 1.

Table 1 provides the various estimates of uninsured Asian people and the associated standard errors for the different small domains for the year 2000. Domain 2 is excluded due to zero sample size. Domain 2 refers to Male Filipinos in the age group 0-17 belonging to MSA's of size 500,000 - 999,999. The measures of precision (posterior s.d.'s) associated with the HB estimates are denoted by $\text{se}(\text{HB})$ and are given by the formula $\text{se}^2(\text{HB}) = \text{var}(\sum_{j=1}^{n_i} w_{ij} p_{ij} \mid \boldsymbol{y})$. One of the advantages of the HB or EB estimates is that for domains with very small sample sizes, often the direct estimates of the proportion of uninsured is zero, whereas the former provide small but non-zero estimates. We chose not to collapse the direct estimates for domains with very small sample sizes. The unit level covariates were quite distinct, and there was no meaningful way to combine them. We note also that when $\lambda = 0.5$, *i.e.*, the direct and synthetic estimates have 1: 2 weight ratio, the EB and HB estimates are very close.

Table 1
Small domain estimates of the proportions of uninsured Asians: Year 2000

Domain	n_i	Direct	'97-'99 average	HB	se (HB)	EB $\lambda = 0.5$	EB $\lambda = 1$	se (EB) $\lambda = 0.5$	se (EB) $\lambda = 1$
1	10	0.126	0.034	0.133	0.043	0.148	0.158	0.057	0.060
2	0	-	0.085	-	-	-	-	-	-
3	24	0.063	0.016	0.074	0.025	0.076	0.082	0.037	0.039
4	28	0.146	0.105	0.150	0.027	0.163	0.171	0.041	0.043
5	20	0.138	0.265	0.143	0.032	0.153	0.160	0.043	0.046
6	17	0.112	0.124	0.120	0.032	0.134	0.144	0.019	0.021
7	78	0.097	0.100	0.104	0.015	0.107	0.112	0.022	0.024
8	66	0.274	0.229	0.253	0.023	0.240	0.224	0.072	0.076
9	5	0.173	0.000	0.164	0.061	0.160	0.154	0.078	0.082
10	6	0.000	0.000	0.033	0.051	0.082	0.123	0.070	0.074
11	7	0.000	0.084	0.032	0.047	0.090	0.134	0.054	0.057
12	11	0.335	0.000	0.302	0.056	0.275	0.245	0.060	0.064
13	7	0.134	0.061	0.134	0.045	0.130	0.128	0.103	0.110
14	2	0.000	0.151	0.020	0.064	0.026	0.039	0.031	0.033
15	27	0.000	0.104	0.023	0.023	0.035	0.052	0.032	0.034
16	29	0.113	0.191	0.119	0.024	0.123	0.127	0.033	0.035
17	27	0.120	0.223	0.127	0.025	0.141	0.152	0.044	0.047
18	14	0.000	0.106	0.024	0.030	0.041	0.062	0.019	0.021
19	77	0.131	0.111	0.133	0.015	0.133	0.134	0.021	0.023
20	75	0.223	0.222	0.213	0.018	0.207	0.200	0.089	0.095
21	3	0.000	0.000	0.022	0.056	0.028	0.043	0.070	0.074
22	6	0.000	0.184	0.026	0.045	0.052	0.079	0.071	0.075
23	8	0.000	0.022	0.037	0.050	0.108	0.162	0.063	0.067
24	9	0.000	0.000	0.029	0.042	0.062	0.093	0.052	0.055
25	10	0.000	0.083	0.023	0.034	0.031	0.046	0.061	0.065
26	6	0.000	0.018	0.020	0.039	0.029	0.044	0.031	0.033
27	32	0.098	0.041	0.105	0.023	0.108	0.114	0.035	0.037
28	23	0.000	0.092	0.024	0.025	0.037	0.055	0.032	0.034
29	25	0.187	0.211	0.173	0.030	0.151	0.134	0.035	0.037
30	23	0.227	0.076	0.210	0.032	0.188	0.169	0.021	0.022

Table 1 (continued)
Small domain estimates of the proportions of uninsured Asians: Year 2000

Domain	n_i	Direct	'97-'99 average	HB	se (HB)	EB $\lambda = 0.5$	EB $\lambda = 1$	se (EB) $\lambda = 0.5$	se (EB) $\lambda = 1$
31	71	0.118	0.059	0.123	0.016	0.125	0.128	0.024	0.026
32	50	0.109	0.156	0.113	0.019	0.112	0.113	0.113	0.120
33	2	0.000	0.000	0.024	0.071	0.037	0.055	0.115	0.122
34	2	0.000	0.000	0.026	0.073	0.047	0.070	0.058	0.061
35	8	0.108	0.000	0.113	0.042	0.112	0.114	0.067	0.071
36	7	0.000	0.000	0.030	0.045	0.065	0.098	0.051	0.054
37	9	0.062	0.197	0.069	0.035	0.062	0.063	0.036	0.038
38	17	0.000	0.019	0.019	0.024	0.023	0.034	0.037	0.040
39	24	0.117	0.022	0.124	0.028	0.134	0.142	0.040	0.043
40	20	0.000	0.070	0.028	0.029	0.052	0.078	0.025	0.027
41	50	0.163	0.145	0.160	0.020	0.156	0.153	0.027	0.029
42	38	0.141	0.114	0.139	0.021	0.133	0.130	0.020	0.022
43	76	0.104	0.090	0.112	0.016	0.120	0.128	0.020	0.022
44	73	0.142	0.149	0.142	0.016	0.139	0.137	0.119	0.127
45	2	0.000	0.000	0.027	0.076	0.051	0.076	0.090	0.095
46	3	0.000	0.052	0.021	0.056	0.023	0.035	0.052	0.055
47	10	0.000	0.072	0.024	0.034	0.044	0.066	0.068	0.072
48	7	0.000	0.172	0.029	0.045	0.068	0.102	0.051	0.054
49	10	0.087	0.364	0.095	0.037	0.099	0.105	0.078	0.083
50	5	0.000	0.000	0.027	0.050	0.053	0.080	0.032	0.034
51	23	0.038	0.092	0.053	0.023	0.056	0.066	0.037	0.039
52	21	0.243	0.195	0.223	0.037	0.198	0.176	0.030	0.032
53	31	0.114	0.184	0.120	0.022	0.121	0.124	0.040	0.042
54	18	0.202	0.169	0.195	0.031	0.188	0.182	0.019	0.020
55	74	0.094	0.115	0.102	0.015	0.102	0.106	0.019	0.020
56	83	0.204	0.296	0.192	0.017	0.178	0.165	0.133	0.141
57	2	0.000	0.124	0.029	0.082	0.062	0.092	0.146	0.154
58	1	0.000	0.000	0.019	0.087	0.023	0.035	0.000	0.000
59	2	0.000	0.196	0.020	0.063	0.021	0.032	0.103	0.194
60	8	0.112	0.116	0.120	0.044	0.132	0.143	0.059	0.063
61	16	0.202	0.140	0.187	0.036	0.169	0.152	0.040	0.043
62	3	0.301	0.163	0.276	0.086	0.252	0.227	0.100	0.107
63	33	0.055	0.093	0.069	0.020	0.073	0.082	0.028	0.030
64	28	0.105	0.275	0.112	0.024	0.115	0.120	0.032	0.034
65	33	0.126	0.133	0.129	0.021	0.126	0.126	0.029	0.031
66	13	0.393	0.290	0.350	0.054	0.323	0.288	0.048	0.051
67	70	0.080	0.136	0.089	0.015	0.088	0.093	0.019	0.021
68	75	0.179	0.233	0.171	0.017	0.159	0.149	0.019	0.021
69	1	0.000	0.000	0.851	0.248	0.705	0.558	0.163	0.173
70	2	0.361	0.000	0.331	0.098	0.299	0.268	0.119	0.126
71	4	0.000	0.091	0.023	0.050	0.032	0.048	0.077	0.082
72	2	0.000	0.182	0.045	0.101	0.157	0.236	0.155	0.165
73	45	0.271	0.144	0.256	0.026	0.256	0.249	0.028	0.030
74	10	0.000	0.044	0.024	0.034	0.034	0.051	0.051	0.055
75	83	0.149	0.097	0.150	0.016	0.160	0.166	0.020	0.021
76	59	0.113	0.205	0.120	0.018	0.128	0.136	0.023	0.024
77	68	0.338	0.224	0.313	0.025	0.302	0.284	0.023	0.024
78	39	0.098	0.138	0.103	0.020	0.102	0.104	0.026	0.028
79	122	0.110	0.163	0.117	0.013	0.125	0.133	0.016	0.017
80	125	0.308	0.314	0.281	0.020	0.262	0.239	0.016	0.017
81	7	0.000	0.000	0.029	0.043	0.066	0.099	0.065	0.069
82	12	0.000	0.045	0.025	0.032	0.047	0.070	0.048	0.051
83	13	0.049	0.017	0.068	0.035	0.088	0.108	0.050	0.053
84	4	0.000	0.061	0.028	0.056	0.060	0.091	0.088	0.093
85	32	0.189	0.113	0.193	0.027	0.217	0.231	0.035	0.037
86	10	0.136	0.056	0.137	0.036	0.127	0.123	0.051	0.054
87	52	0.192	0.098	0.185	0.021	0.184	0.180	0.024	0.026
88	65	0.153	0.120	0.155	0.018	0.162	0.166	0.022	0.024
89	71	0.285	0.210	0.265	0.022	0.256	0.242	0.022	0.023
90	57	0.086	0.146	0.095	0.017	0.102	0.110	0.022	0.024
91	153	0.149	0.167	0.150	0.011	0.156	0.160	0.014	0.015
92	138	0.308	0.285	0.283	0.020	0.266	0.244	0.015	0.017
93	10	0.000	0.000	0.030	0.041	0.073	0.110	0.059	0.063
94	16	0.067	0.015	0.081	0.029	0.090	0.101	0.042	0.044
95	18	0.108	0.018	0.123	0.032	0.145	0.163	0.046	0.049
96	14	0.111	0.087	0.125	0.039	0.160	0.185	0.050	0.053

The HB estimates of the proportion of uninsured for Asians vary in the 2%-35% range for the different small domains excluding domain 69. Admittedly, the EB and HB estimates for domain 69 are very adversely affected due to small sample size. We also report the standard errors associated with the HB estimates, and estimated approximate root mean squares accompanying the EB estimates. The proposed approach largely overcomes the valid criticism that naive EB estimates of standard errors (which ignore the $O(k^{-1})$ term) are typically underestimates. We have also provided a column giving the 3-year average of the direct estimates in 1997-1999. This is primarily to examine whether domains with zero direct estimates in 2000 also possess the same feature in other years, and also for comparison of EB and HB estimates with these estimates rather than the direct estimates. It turns out that with very few exceptions, the 1997-1999 average do not conform very much to the direct estimates. However, domain 69 still has zero direct estimate.

Table 2 provides the summary table for the proprtion of uninsured for the three age groups 0-17, 18-64 and 65+ individually for Chinese (Asian 1), Filipino (Asian 2), Asian Indian (Asian 3) and other Asians (Asian 4). It turns out that at this higher level of aggregation, both the EB and HB small domain estimates are fairly close to the corresponding direct estimates except possibly for the age-group 65+. This seems to be quite satisfactory, since at this level of aggregation, the direct estimates often serve as benchmarks for comparison purpose.

Table 2
Proportions without health insurance coverage by age group and Asian group in 2000

	Direct	HB	EB ($\lambda = 0.5$)	EB ($\lambda = 1$)
0-17 years				
Total	0.120	0.126	0.131	0.137
Asian 1	0.087	0.097	0.105	0.114
Asian 2	0.046	0.062	0.071	0.083
Asian 3	0.113	0.117	0.114	0.114
Asian 4	0.165	0.165	0.171	0.175
18-64 years				
Total	0.177	0.172	0.168	0.164
Asian 1	0.162	0.160	0.160	0.159
Asian 2	0.137	0.137	0.135	0.134
Asian 3	0.150	0.147	0.141	0.137
Asian 4	0.219	0.208	0.203	0.195
65+ years				
Total	0.063	0.080	0.103	0.123
Asian 1	0.083	0.097	0.123	0.143
Asian 2	0.021	0.043	0.064	0.085
Asian 3	0.119	0.126	0.136	0.145
Asian 4	0.055	0.075	0.100	0.123

6. Model diagnostics and implementation of the hierarchical Bayesian model

We followed Gelman and Rubin (1992) for the implementation and convergence diagnostics of the Gibbs sampler. In particular we took 5 chains each of size 1,000 with an initial burning period of 1,000 iterations. We checked the potential scale reduction factors for convergence and these appeared to be very close to unity (= 1 at convergence) for each one of the parameters. A number of other diagnostics criteria are available in the literature, and are implemented via the software CODA. A partial output is provided in the Figure 1. The left side shows the overlap of the 5 parallel chains, and the right side shows the posterior inference for each parameter and the deviance (-2 log likelihood). For details regarding the description of the software that we used, we refer to Appendix C of Gelman, Carlin, Stern and Rubin (2004).

A Bayesian way to check the fit of a model to data is to draw simulated values from the posterior predictive distribution of replicated data and compare these samples to observed data. A wide departure between the generated and the observed data indicates lack of fit of the model. Following Gelman *et al.* (2004), we calculated the Bayesian p-values for checking the goodness-of-fit of the proposed Bayesian models. The general rationale behind such calculations is as follows. Let y denote the vector of observed data, ξ the vector of unknown parameters, $f(y \mid \xi)$ the density of y given ξ and $\Pi(\xi \mid y)$, the posterior density of ξ given y. Suppose one has drawn samples $\xi^{(1)}, \ldots, \xi^{(R)}$ from this posterior distribution using MCMC simulation. Simulate now R hypothetical replicates of the data, say $y^{(1)}, \ldots, y^{(R)}$, where $y^{(l)}, (l = 1, \ldots, R)$ is drawn from the conditional distribution of y given the simulated $\xi^{(l)}$. If the model is reasonably accurate, these hypothetical replicates should be similar to the observed data y. This is formally done by first choosing a divergence variable, say $d(y, \xi)$ which will have an extreme value if the data y are in complete disagreement with the given model. Then a p-value is estimated by the proportion of cases in which the simulated divergence variable exceeds the realized value of the same. Thus the estimated p-value (usually referred to as the posterior predictive p-value) is equal to $R^{-1} \sum_{l=1}^{R} I_{[d(y^{(l)}, \xi^{(l)}) \geq d(y, \xi^{(l)})]}$, where I is the usual indicator function. One way of checking the goodness of fit of the model is by a scatter plot of realized values $d(y, \xi^{(l)})$ against the predictive values $d(y^{(l)}, \xi^{(l)})$ on the same scale. A good fit is indicated by about half the points in the scatter plot falling above the 45^0 line, and half falling below. In other words, for large samples, the estimated p-value will not be far away from one half. Of course, one may also carry out a graphical analysis by using different plots for different subgroups, thereby allowing visualization of possible local model failure which may otherwise be obscured in the aggregate plot.

There are several possible choices of the divergence variable d. We considered a particular one in the present case. Noting that $E(Y_{ij} \mid p_{ij}) = p_{ij} = \exp(\theta_{ij})/(1+\exp(\theta_{ij}))$, one can consider the squared standardized residuals $((y_{ij} - p_{ij}^{(l)})^2)/(p_{ij}^{(l)}(1-p_{ij}^{(l)}))$, where $p_{ij}^{(l)} = \exp(\theta_{ij}^{(l)})/(1+\exp(\theta_{ij}^{(l)}))$ are the generated values of the p_{ij} from the l^{th} iteration. Then the divergence variable d is

$$d(y, p^{(l)}) = \sum_{i=1}^{95} \sum_{j=1}^{n_i} \frac{(y_{ij} - p_{ij}^{(l)})^2}{p_{ij}^{(l)}(1-p_{ij}^{(l)})}$$

$$d(y^{(l)}, p^{(l)}) = \sum_{i=1}^{95} \sum_{j=1}^{n_i} \frac{(y_{ij}^{(l)} - p_{ij}^{(l)})^2}{p_{ij}^{(l)}(1-p_{ij}^{(l)})}.$$

Clearly, there are other possible choices of d. Gelfand and Ghosh (1998) proposed a number of divergence measures, and studied their properties.

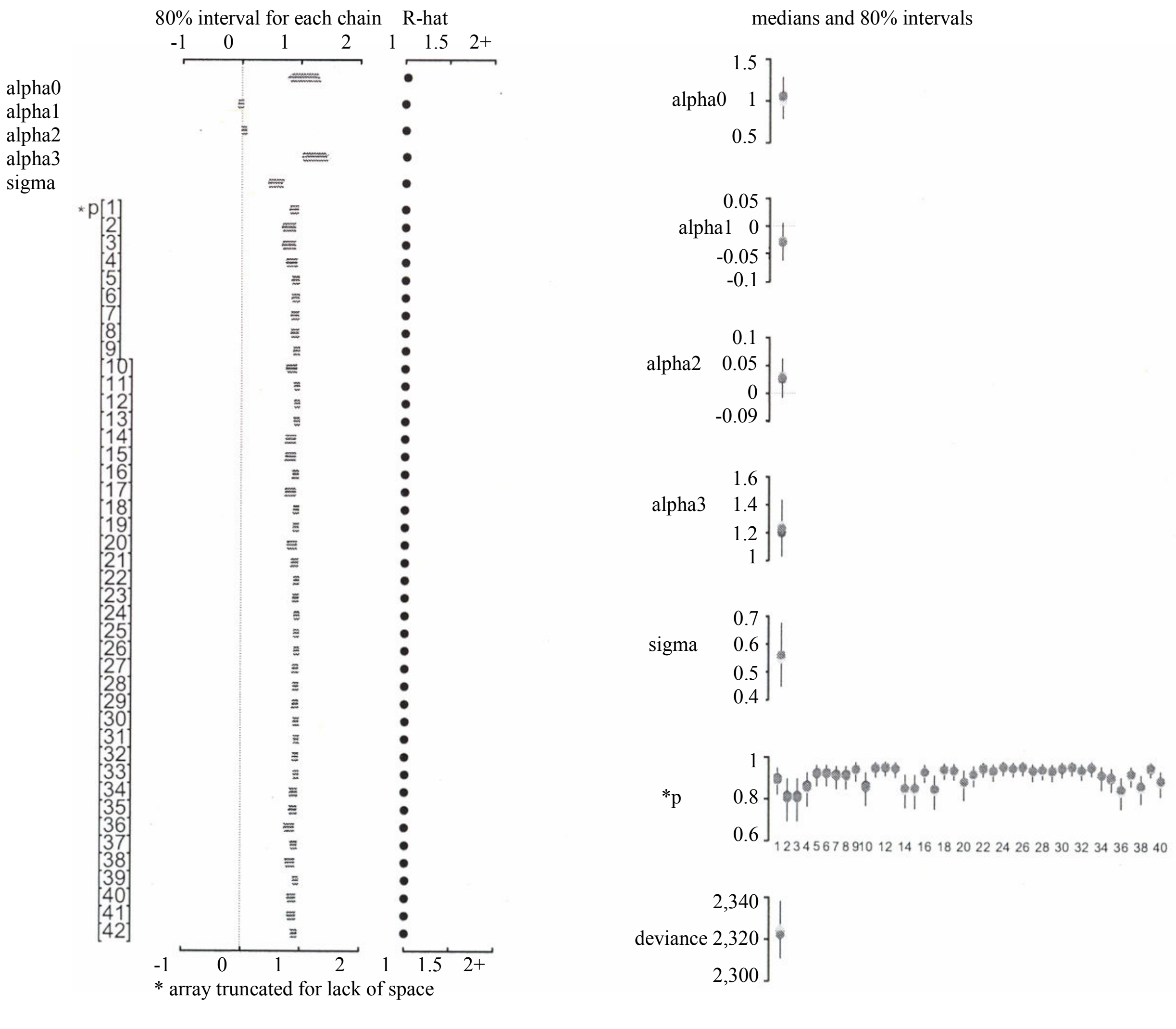

Figure 1 Bugs model at "asian_model.bug", 5 chains, each with 1,000 iterations

For the hierarchical Bayesian logistic regression model, the estimated p-value is 0.4216 for $(c, d) = (0.02, 0.02)$. The other choices of (c, d) produce similar values. The p-value bigger than 0.3 is usually treated as a good fit. Thus the proposed HB procedure seems to work well in this situation.

We have also calculated the $p_D =$ var(deviance)/2 and the *deviance information criterion* DIC or the estimated predictive deviance. The p_D can be thought of as the number of 'unconstrained parameters in the model, where a parameter is counted as 1 if it is a part of the original model (data distribution) and is 0 if it is associated with any prior distribution. The DIC is estimated as

$$\text{DIC} = 2\hat{D}(y, \theta^{(l)}) - \hat{D}(y, \hat{\theta})$$

where $\hat{D}(y, \hat{\theta})$ is the deviance calulated at the estimated parameters and $\hat{D}(y, \theta^{(l)})$ is the esimated deviance using posterior simulation. For details, see Gelman *et al.* (2004).

For our HB analysis $p_D = 56.75$ and DIC = 2,414.41. Usually the p_D and DIC are used as criteria of model fitting and to select the model with best predictive power. Thus, we fit also the simple logistic regression model (current model without any random effects) which means that there is no data pooling, and the estimated p_D and DIC are 22.60 and 2,379.55 respectively. The corresponding p-value is 0.3848. Thus the proposed model seems to fit the data reasonably well

7. Summary, future work and discussion

Estimating the proportion of uninsured people, especially among the minorities, is definitely a problem of great importance, and is likely to affect the policy making of Federal and State agencies. We have just started addressing this very important issue, and have provided both empirical and hierarchical Bayesian small domain estimates for the Asian subpopulation cross-classified by age, sex and other demographic characteristics. We have also discussed the adequacy of our model fit via posterior predictive p-value. Much work remains to be done however. In particular, we want to extend the present findings to the analysis of bivariate and multivariate binary data.

As pointed out by a reviewer, the present analysis ignores household clustering in the likelihood, since the original survey was a household survey, and very definitely, insurance coverage is correlated within households. However, we have assumed only a conditionally independent hierarchical model given the covariates and the random effects. Once, we have assigned distributions to the random effects, and subsequently distributions to the regression coefficients and the variance components, dependence is built automatically in the final model, both at the unit and domain levels. Moreover, as mentioned earlier, adequacy of the hierarchical Bayesian model has been tested through posterior predictive p-values.

As a final comment, the research presented here is for illustrative purposes only. Implementation of this method for policy related matters would require further consideratuions of the methods and adherence to institutional standards for official policy release.

Acknowledgements

We thank the AE and two reviewers for their constructive comments on earlier drafts of the paper. The first author's research was partially funded by NSF Grants SES-0317589 and SES-0631426. The fourth author's research was partially supported by an NSF Grant SES-0318184. The research was also partially supported by NCHS/CDC under Project Number 282286285 entitled "Model-Based Subdomain Estimates". The findings and conclusions in this paper are those of the authors and do not necessarily represent the views of the National Center for Health Statistics, Centers for Disease Control and Prevention.

Appendix

Proof of Theorem 1.

$$\text{MSE}(\hat{\mu}_{iw}^{\text{EB}})$$

$$= E(\hat{\mu}_{iw}^{\text{EB}} - \hat{\mu}_{iw})^2 = E\left(\sum_{j=1}^{n_i} w_{ij}(\hat{p}_{ij}^{\text{EB}} - p_{ij})\right)^2$$

$$= E\left(\sum_{j=1}^{n_i} w_{ij}(\hat{p}_{ij}^{\text{EB}} - \hat{p}_{ij}^{B} + \hat{p}_{ij}^{B} - p_{ij})\right)^2$$

$$= E\left(\sum_{j=1}^{n_i} w_{ij}(\hat{p}_{ij}^{\text{EB}} - \hat{p}_{ij}^{B})\right)^2 + E\left(\sum_{j=1}^{n_i} w_{ij}(\hat{p}_{ij}^{B} - p_{ij})\right)^2$$

$$+ 2E\left[\left(\sum_{j=1}^{n_i} w_{ij}(\hat{p}_{ij}^{\text{EB}} - \hat{p}_{ij}^{B})\right)\left(\sum_{j=1}^{n_i} w_{ij}(\hat{p}_{ij}^{B} - p_{ij})\right)\right].$$

Noting that $E(p_{ij} \mid \boldsymbol{y}) = \hat{p}_{ij}^{B}$,

$$E\left[\left(\sum_{j=1}^{n_i} w_{ij}(\hat{p}_{ij}^{\text{EB}} - \hat{p}_{ij}^{B})\right)\left(\sum_{j=1}^{n_i} w_{ij}(\hat{p}_{ij}^{B} - p_{ij})\right)\right]$$

$$= E\left[\left(\sum_{j=1}^{n_i} w_{ij}(\hat{p}_{ij}^{\text{EB}} - \hat{p}_{ij}^{B})\right) \times E\left(\sum_{j=1}^{n_i} w_{ij}(\hat{p}_{ij}^{B} - p_{ij}) \mid \boldsymbol{y}\right)\right] = 0.$$

Hence,

$$\text{MSE}(\hat{\mu}_{iw}^{\text{EB}}) = E\left(\sum_{j=1}^{n_i} w_{ij}(\hat{p}_{ij}^{\text{EB}} - \hat{p}_{ij}^{B})\right)^2 + E\left(\sum_{j=1}^{n_i} w_{ij}(\hat{p}_{ij}^{B} - p_{ij})\right)^2. \tag{A.1}$$

But

$$E\left(\sum_{j=1}^{n_i} w_{ij}(\hat{p}_{ij}^{B} - p_{ij})\right)^2 = \sum_{j=1}^{n_i} w_{ij}^2 E(\hat{p}_{ij}^{B} - p_{ij})^2.$$

Next we calculate

$$\begin{aligned} &E(\hat{p}_{ij}^{B} - p_{ij})^2 \\ &= E\left[\frac{1}{\lambda+1} y_{ij} + \frac{\lambda}{\lambda+1} m_{ij}(\boldsymbol{b}) - p_{ij}\right]^2 \\ &= E\left[\frac{1}{\lambda+1}(y_{ij} - p_{ij}) + \frac{\lambda}{\lambda+1}(m_{ij}(\boldsymbol{b}) - p_{ij})\right]^2 \\ &= \frac{1}{(\lambda+1)^2} E(y_{ij} - p_{ij})^2 + \frac{\lambda^2}{(\lambda+1)^2} E\left(m_{ij}(\boldsymbol{b}) - p_{ij}\right)^2 \\ &\quad + \frac{2\lambda}{(\lambda+1)^2} E(y_{ij} - p_{ij})(m_{ij}(\boldsymbol{b}) - p_{ij}) \\ &= \frac{1}{(\lambda+1)^2} E(p_{ij}(1 - p_{ij})) + \frac{\lambda^2}{(\lambda+1)^2} V(p_{ij}) + 0 \\ &= \frac{1}{(\lambda+1)^2}\left(\frac{\lambda}{(\lambda+1)} m_{ij}(\boldsymbol{b})(1 - m_{ij}(\boldsymbol{b}))\right) \\ &\quad + \frac{\lambda^2}{(\lambda+1)^2}\left(\frac{m_{ij}(\boldsymbol{b})(1 - m_{ij}(\boldsymbol{b}))}{\lambda+1}\right) \\ &= \frac{\lambda m_{ij}(\boldsymbol{b})(1 - m_{ij}(\boldsymbol{b}))}{(\lambda+1)^2}, \end{aligned}$$

so that

$$E\left[\sum_{j=1}^{n_i} w_{ij}(\hat{p}_{ij}^{B} - p_{ij})\right]^2 = \frac{\lambda}{(\lambda+1)^2} \sum_{j=1}^{n_i} w_{ij}^2 m_{ij}(\boldsymbol{b})(1 - m_{ij}(\boldsymbol{b})). \tag{A.2}$$

Finally, we calculate,

$$\begin{aligned} &E\left[\sum_{j=1}^{n_i} w_{ij}(\hat{p}_{ij}^{\text{EB}} - \hat{p}_{ij}^{B})\right]^2 \\ &= \frac{\lambda^2}{(\lambda+1)^2} E\left[\sum_{j=1}^{n_i} w_{ij}(m_{ij}(\hat{\boldsymbol{b}}) - m_{ij}(\boldsymbol{b}))\right]^2 \\ &= \frac{\lambda^2}{(\lambda+1)^2} E\left[\sum_{j=1}^{n_i} w_{ij}^2 (m_{ij}(\hat{\boldsymbol{b}}) - m_{ij}(\boldsymbol{b}))^2 \right. \\ &\quad + \sum\sum_{1 \le j \ne j' \le n_i} w_{ij} w_{ij'} (m_{ij}(\hat{\boldsymbol{b}}) - m_{ij}(\boldsymbol{b})) \\ &\quad \left. (m_{ij'}(\hat{\boldsymbol{b}}) - m_{ij'}(\boldsymbol{b}))\right]. \end{aligned} \tag{A.3}$$

By two-step Taylor expansion,

$$m_{ij}(\hat{\boldsymbol{b}}) \doteq m_{ij}(\boldsymbol{b}) + \left(\frac{\partial m_{ij}(\boldsymbol{b})}{\partial \boldsymbol{b}}\right)^T (\hat{\boldsymbol{b}} - \boldsymbol{b}) + \frac{1}{2}(\hat{\boldsymbol{b}} - \boldsymbol{b})^T \frac{\partial^2 m_{ij}(\boldsymbol{b})}{\partial \boldsymbol{b} \partial \boldsymbol{b}^T}(\hat{\boldsymbol{b}} - \boldsymbol{b}).$$

Noting that $(\partial^2 m_{ij}(\boldsymbol{b}))/(\partial \boldsymbol{b} \partial \boldsymbol{b}^T) = (1 - 2m_{ij}(\boldsymbol{b}))m_{ij}(\boldsymbol{b})(1 - m_{ij}(\boldsymbol{b}))\boldsymbol{x}_{ij}\boldsymbol{x}_{ij}^T$, it follows that

$$\begin{aligned} &E[m_{ij}(\hat{\boldsymbol{b}}) - m_{ij}(\boldsymbol{b})]^2 \\ &\doteq E\left[m_{ij}(\boldsymbol{b})(1 - m_{ij}(\boldsymbol{b}))\boldsymbol{x}_{ij}^T(\hat{\boldsymbol{b}} - \boldsymbol{b}) + \frac{1}{2}(\hat{\boldsymbol{b}} - \boldsymbol{b})^T \right. \\ &\quad \left. m_{ij}(\boldsymbol{b})(1 - m_{ij}(\boldsymbol{b}))(1 - 2m_{ij}(\boldsymbol{b}))\boldsymbol{x}_{ij}\boldsymbol{x}_{ij}^T(\hat{\boldsymbol{b}} - \boldsymbol{b})\right]^2 \\ &= m_{ij}^2(\boldsymbol{b})(1 - m_{ij}(\boldsymbol{b}))^2 E\left[\boldsymbol{x}_{ij}^T(\hat{\boldsymbol{b}} - \boldsymbol{b}) + \frac{1}{2}(1 - 2m_{ij}(\boldsymbol{b})) \right. \\ &\quad \left. (\hat{\boldsymbol{b}} - \boldsymbol{b})^T \boldsymbol{x}_{ij}\boldsymbol{x}_{ij}^T(\hat{\boldsymbol{b}} - \boldsymbol{b})\right]^2. \end{aligned} \tag{A.4}$$

The first neglected term is $O_p(\|\hat{\boldsymbol{b}} - \boldsymbol{b}\|^3)$. From Sarkar and Ghosh (1998), $\hat{\boldsymbol{b}} - \boldsymbol{b}$ is asymptotically $\text{N}(0, \boldsymbol{\Sigma}^{-1}(\boldsymbol{b}))$, where $\boldsymbol{\Sigma}(\boldsymbol{b})$ is defined before Theorem 1. With the assumption that $\boldsymbol{\Sigma}(\boldsymbol{b}) = O_e(k)$, it follows that $\boldsymbol{\Sigma}^{-1}(\boldsymbol{b}) = O_e(k^{-1})$. Thus, $\|\hat{\boldsymbol{b}} - \boldsymbol{b}\| = O_p(k^{-1/2})$. Hence, the first neglected term is $O_p(k^{-3/2})$. Next, we observe that

$$\begin{aligned} E[\boldsymbol{x}_{ij}^T(\hat{\boldsymbol{b}} - \boldsymbol{b})]^2 &= E[(\hat{\boldsymbol{b}} - \boldsymbol{b})^T \boldsymbol{x}_{ij}\boldsymbol{x}_{ij}^T(\hat{\boldsymbol{b}} - \boldsymbol{b})] \\ &= tr[\boldsymbol{x}_{ij}\boldsymbol{x}_{ij}^T E(\hat{\boldsymbol{b}} - \boldsymbol{b})(\hat{\boldsymbol{b}} - \boldsymbol{b})^T]. \end{aligned} \tag{A.5}$$

In order to find $E[(\hat{\boldsymbol{b}}-\boldsymbol{b})(\hat{\boldsymbol{b}}-\boldsymbol{b})^T]$, we proceed as follows: Let $\boldsymbol{T}(\boldsymbol{b})=\sum_{i=1}^k\sum_{j=1}^{n_i}(y_{ij}-m_{ij}(\boldsymbol{b}))\boldsymbol{x}_{ij}$ so that $\boldsymbol{T}(\hat{\boldsymbol{b}})=\boldsymbol{0}$. By one-step Taylor expansion, $\boldsymbol{0}=\boldsymbol{T}(\hat{\boldsymbol{b}})=\boldsymbol{T}(\boldsymbol{b})+[\nabla\boldsymbol{T}(\boldsymbol{b})]^T(\hat{\boldsymbol{b}}-\boldsymbol{b})+O_p(n_T^{-1})$, where

$$\nabla\boldsymbol{T}(\boldsymbol{b})=-\sum_{i=1}^k\sum_{j=1}^{n_i}\left(\frac{\partial m_{ij}(\boldsymbol{b})}{\partial\boldsymbol{b}}\right)\boldsymbol{x}_{ij}^T$$

$$=-\sum_{i=1}^k\sum_{j=1}^{n_i}m_{ij}(\boldsymbol{b})(1-m_{ij}(\boldsymbol{b}))\boldsymbol{x}_{ij}\boldsymbol{x}_{ij}^T$$

$$=-\boldsymbol{X}^T\boldsymbol{M}(\boldsymbol{I}-\boldsymbol{M})\boldsymbol{X}$$

$$=-\boldsymbol{\Sigma}(\boldsymbol{b}). \tag{A.6}$$

Thus, $\hat{\boldsymbol{b}}-\boldsymbol{b}=\boldsymbol{\Sigma}^{-1}\boldsymbol{T}(\boldsymbol{b})+O_p(n_T^{-1})$. Since $V(y_{ij})=m_{ij}(\boldsymbol{b})(1-m_{ij}(\boldsymbol{b}))$, $V(T(\boldsymbol{b}))=\boldsymbol{\Sigma}(\boldsymbol{b})$. Hence $E[(\hat{\boldsymbol{b}}-\boldsymbol{b})(\hat{\boldsymbol{b}}-\boldsymbol{b})^T]=\boldsymbol{\Sigma}^{-1}(\boldsymbol{b})+O(n_T^{-3/2})$. Also, in (8.5), we have $E(\boldsymbol{x}_{ij}^T)(\hat{\boldsymbol{b}}-\boldsymbol{b})^2=tr(\boldsymbol{x}_{ij}\boldsymbol{x}_{ij}^T\boldsymbol{\Sigma}^{-1}(\boldsymbol{b}))+O_p(n_T^{-1})$. Accordingly, by (A.4) and (A.5), we have the approximation

$$E[m_{ij}(\hat{\boldsymbol{b}})-m_{ij}(\boldsymbol{b})]^2=m_{ij}^2(\boldsymbol{b})(1-m_{ij}(\boldsymbol{b}))^2\boldsymbol{x}_{ij}^T\boldsymbol{\Sigma}^{-1}(\boldsymbol{b})\boldsymbol{x}_{ij} \tag{A.7}$$

which is correct up to $O(n_T^{-1})$ by our assumption. Note that the neglected term

$$E[\boldsymbol{x}_{ij}^T(\hat{\boldsymbol{b}}-\boldsymbol{b})(1-2m_{ij}(\boldsymbol{b}))(\hat{\boldsymbol{b}}-\boldsymbol{b})^T\boldsymbol{x}_{ij}\boldsymbol{x}_{ij}^T(\hat{\boldsymbol{b}}-\boldsymbol{b})]$$

$$=O(n_T^{-3/2})$$

since

$$E[\boldsymbol{x}_{ij}^T(\hat{\boldsymbol{b}}-\boldsymbol{b})(1-2m_{ij}(\boldsymbol{b}))(\hat{\boldsymbol{b}}-\boldsymbol{b})^T\boldsymbol{x}_{ij}\boldsymbol{x}_{ij}^T(\hat{\boldsymbol{b}}-\boldsymbol{b})]$$

$$=(1-2m_{ij}(\boldsymbol{b}))E[\boldsymbol{x}_{ij}^T(\hat{\boldsymbol{b}}-\boldsymbol{b})(\hat{\boldsymbol{b}}-\boldsymbol{b})^T\boldsymbol{x}_{ij}]$$

$$=(1-2m_{ij}(\boldsymbol{b}))E[\boldsymbol{x}_{ij}^T(\hat{\boldsymbol{b}}-\boldsymbol{b})]^3$$

$$=O(n_T^{-3/2}).$$

Similarly, note that $E[\boldsymbol{x}_{ij}^T(\hat{\boldsymbol{b}}-\boldsymbol{b})]^4=O(n_T^{-2})$ and

$$E[(m_{ij}(\hat{\boldsymbol{b}})-m_{ij}(\boldsymbol{b}))(m_{ij'}(\hat{\boldsymbol{b}})-m_{ij'}(\boldsymbol{b}))]$$

$$=m_{ij}(\boldsymbol{b})(1-m_{ij}(\boldsymbol{b}))m_{ij'}(\boldsymbol{b})(1-m_{ij'}(\boldsymbol{b}))\boldsymbol{x}_{ij}^T\boldsymbol{\Sigma}^{-1}(\boldsymbol{b})\boldsymbol{x}_{ij'}$$

$$+O(n_T^{-3/2}). \tag{A.8}$$

This leads to

$$E\left[\sum_{j=1}^{n_i}w_{ij}(\hat{p}_{ij}^{\text{EB}}-\hat{p}_{ij}^B)\right]^2$$

$$=\frac{\lambda^2}{(\lambda+1)^2}\left[\sum_{j=1}^{n_i}w_{ij}^2m_{ij}^2(\boldsymbol{b})(1-m_{ij}(\boldsymbol{b}))^2\boldsymbol{x}_{ij}^T\boldsymbol{\Sigma}^{-1}(\boldsymbol{b})\boldsymbol{x}_{ij}\right.$$

$$+\sum\sum_{1\le j\ne j'\le n_i}w_{ij}w_{ij'}m_{ij}(\boldsymbol{b})(1-m_{ij}(\boldsymbol{b}))m_{ij'}(\boldsymbol{b})$$

$$\left.(1-m_{ij'}(\boldsymbol{b}))\boldsymbol{x}_{ij}^T\boldsymbol{\Sigma}^{-1}(\boldsymbol{b})\boldsymbol{x}_{ij'}\right]+O(n_T^{-3/2})$$

$$=\frac{\lambda^2}{(\lambda+1)^2}\left[\sum_{j=1}^{n_i}w_{ij}m_{ij}(\boldsymbol{b})(1-m_{ij}(\boldsymbol{b}))\boldsymbol{x}_{ij}\right]^T$$

$$\times\boldsymbol{\Sigma}^{-1}(\boldsymbol{b})\left[\sum_{j=1}^{n_i}w_{ij}m_{ij}(\boldsymbol{b})(1-m_{ij}(\boldsymbol{b}))\boldsymbol{x}_{ij}\right]+O(n_T^{-3/2}). \tag{A.9}$$

Since $\boldsymbol{\Sigma}^{-1}(\boldsymbol{b})=O(k^{-1})$, and $n_T=O_e(k)$ by our assumption, the theorem follows from (A.2) and (A.9).

Proof of Theorem 2. We first note that $\hat{\boldsymbol{b}}=\boldsymbol{b}+O_p(n_T^{-1/2})=\boldsymbol{b}+O_p(K^{-1})$ and $\boldsymbol{\Sigma}^{-1}(\boldsymbol{b})=O(k^{-1})$. Hence, the second term in the right hand side of (8.7) is approximated by

$$c\left[\sum_{j=1}^{n_i}w_{ij}m_{ij}(\hat{\boldsymbol{b}})(1-m_{ij}(\hat{\boldsymbol{b}}))\boldsymbol{x}_{ij}\right]^T$$

$$\boldsymbol{\Sigma}^{-1}(\hat{\boldsymbol{b}})\left[\sum_{j=1}^{n_i}w_{ij}m_{ij}(\hat{\boldsymbol{b}})(1-m_{ij}(\hat{\boldsymbol{b}}))\boldsymbol{x}_{ij}\right] \tag{A.10}$$

$(c=\lambda^2/(1+\lambda)^2)$ which is correct up to $\mathrm{O}(k^{-1})$. However, if we estimate $m_{ij}(\boldsymbol{b})(1-m_{ij}(\boldsymbol{b}))$ simply by $m_{ij}(\hat{\boldsymbol{b}})(1-m_{ij}(\hat{\boldsymbol{b}}))$, we will be ignoring the $O(k^{-1})$ term. Thus, we need a careful approximation of the bias $E(\hat{\boldsymbol{b}}-\boldsymbol{b})$ to achieve the desired approximation. To this end, we follow Cox and Snell (1968).

We begin with the identity

$$E[m_{ij}(\hat{\boldsymbol{b}})(1-m_{ij}(\hat{\boldsymbol{b}}))]$$

$$=E[(m_{ij}(\boldsymbol{b})+m_{ij}(\hat{\boldsymbol{b}})-m_{ij}(\boldsymbol{b}))(1-m_{ij}(\boldsymbol{b})+m_{ij}(\boldsymbol{b})-m_{ij}(\hat{\boldsymbol{b}}))]$$

$$=m_{ij}(\boldsymbol{b})(1-m_{ij}(\boldsymbol{b}))+(1-2m_{ij}(\boldsymbol{b}))E[m_{ij}(\hat{\boldsymbol{b}})-m_{ij}(\boldsymbol{b})]$$

$$-E[m_{ij}(\hat{\boldsymbol{b}})-m_{ij}(\boldsymbol{b})]^2.$$

Now, again by a two-step Taylor expansion,

$$E[m_{ij}(\hat{\boldsymbol{b}}) - m_{ij}(\boldsymbol{b})] = \left[\frac{\partial m_{ij}(\boldsymbol{b})}{\partial \boldsymbol{b}}\right]^T E(\hat{\boldsymbol{b}} - \boldsymbol{b})$$

$$+ \frac{1}{2} E\left[(\hat{\boldsymbol{b}} - \boldsymbol{b})^T \frac{\partial^2 m_{ij}(\boldsymbol{b})}{\partial \boldsymbol{b} \partial \boldsymbol{b}^T} (\hat{\boldsymbol{b}} - \boldsymbol{b})\right] + O(n_T^{-3/2}).$$

In order to find $E(\hat{\boldsymbol{b}} - \boldsymbol{b})$, we proceed as follows. We begin with the second order Taylor expansion

$$0 = T_r(\hat{\boldsymbol{b}}) = T_r(\boldsymbol{b}) + \sum_{s=1}^{p} (\hat{b}_s - b_s) \frac{\partial T_r(\boldsymbol{b})}{\partial b_s}$$

$$+ \frac{1}{2} \sum_{s=1}^{p} \sum_{t=1}^{p} (\hat{b}_s - b_s)(\hat{b}_t - b_t) \frac{\partial^2 T_r(\boldsymbol{b})}{\partial b_s \partial b_t} + O(n_T^{-3/2}).$$

Taking expectations and following Cox and Snell (1968),

$$0 = E(T_r(\hat{\boldsymbol{b}}))$$

$$= \sum_{s=1}^{p} \left[E(\hat{b}_s - b_s) E\left(\frac{\partial T_r(\boldsymbol{b})}{\partial b_s}\right) + \text{Cov}\left(\hat{b}_s - b_s, \frac{\partial T_r(\boldsymbol{b})}{\partial b_s}\right)\right]$$

$$+ \frac{1}{2} \sum_{s=1}^{p} \sum_{t=1}^{p} (E(\hat{b}_s - b_s)(\hat{b}_t - b_t)) \left(\frac{\partial^2 T_r(\boldsymbol{b})}{\partial b_s \partial b_t}\right)$$

$$+ \frac{1}{2} \sum_{s=1}^{p} \sum_{t=1}^{p} \text{Cov}\left[(\hat{b}_s - b_s)(\hat{b}_t - b_t), \left(\frac{\partial^2 T_r(\boldsymbol{b})}{\partial b_s \partial b_t}\right)\right]$$

$$+ O(n_T^{-3/2}) = -\sum_{s=1}^{p} E(\hat{b}_s - b_s) \sigma_{rs}$$

$$+ \sum_{s=1}^{p} \sum_{u=1}^{p} \text{Cov}\left[\sigma^{su}(\boldsymbol{b}) T_u(\boldsymbol{b}), \frac{\partial T_r(\boldsymbol{b})}{\partial b_s}\right]$$

$$+ \frac{1}{2} \sum_{s=1}^{p} \sum_{t=1}^{p} \sigma^{st} E\left[\frac{\partial^2 T_r(\boldsymbol{b})}{\partial b_s \partial b_t}\right] + O(n_T^{-3/2}). \quad \text{(A.11)}$$

Note $\text{Cov}[\sigma^{su}(\boldsymbol{b}) T_u(\boldsymbol{b}), \partial T_r(\boldsymbol{b}) / \partial b_s] = 0$ since $\partial T_r(\boldsymbol{b}) / \partial b_s$ is a constant independent of the y_{ij}.
Similarly,

$$\text{Cov}\left[(\hat{b}_s - b_s)(\hat{b}_t - b_t), \left(\frac{\partial^2 T_r(\boldsymbol{b})}{\partial b_s \partial b_t}\right)\right] = 0.$$

Also, let

$$K_{rst} = E\left[\frac{\partial^2 T_r(\boldsymbol{b})}{\partial b_s \partial b_t}\right]$$

$$= \frac{\partial}{\partial b_t} \sum_{i=1}^{k} \sum_{j=1}^{n_i} - m_{ij}(\boldsymbol{b})(1 - m_{ij}(\boldsymbol{b})) x_{ijr}\, x_{ijs}$$

$$= -\sum_{i=1}^{k} \sum_{j=1}^{n_i} (1 - 2m_{ij}(\boldsymbol{b})) m_{ij}(\boldsymbol{b})(1 - m_{ij}(\boldsymbol{b})) x_{ijr}\, x_{ijs}\, x_{ijt}. \quad \text{(A.12)}$$

Thus, one has

$$\sum_{s=1}^{k} \sigma_{rs} E(\hat{b}_s - b_s) \doteq \sum_{s=1}^{k} \sum_{t=1}^{p} \sigma^{su} K_{rst},\ r = 1, \ldots, p.$$

In matrix notations, one gets

$$\boldsymbol{\Sigma} E(\hat{\boldsymbol{b}} - \boldsymbol{b}) = \frac{1}{2} \begin{pmatrix} tr(\boldsymbol{\Sigma}^{-1} \boldsymbol{K}_1) \\ . \\ . \\ . \\ tr(\boldsymbol{\Sigma}^{-1} \boldsymbol{K}_p) \end{pmatrix}$$

where $\boldsymbol{K}_r = ((K_{rst}))$.

Hence,

$$E(\hat{\boldsymbol{b}} - \boldsymbol{b}) \doteq \frac{1}{2} \boldsymbol{\Sigma}^{-1} \begin{pmatrix} tr(\boldsymbol{\Sigma}^{-1} \boldsymbol{K}_1) \\ . \\ . \\ . \\ tr(\boldsymbol{\Sigma}^{-1} \boldsymbol{K}_p) \end{pmatrix} + O(n_T^{-3/2}).$$

Since $n_T = O_e(k)$, the theorem follows.

References

Botman, S.L., Moore, T.F., Moriarity, C.L. and Parsons, V.L. (2000). Design and estimation for the National Health Interview Survey, 1995-2004. *Vital and Health Statistics*, 2, 130.

Cox, D.R., and Snell, E.J. (1968). A general distribution of residuals (with discussion). *Journal of the Royal Statistical Society*, Series B, 30, 248-275.

Ghosh, M., and Maiti, T. (2004). Small-area estimation based on natural exponential family quadratic variance function models and survey weights. *Biometrika*, 91, 95-112.

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.

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

Gelman, A., and Rubin, D. (1992). Inference from iterative simulation using multiple sequences (with discussion). *Statistical Science*, 457-511.

Godambe, V.P., and Thompson, M.E. (1989). An extension of quasi-likelihood estimation (with discussion). *Journal of Statistical Planning and Infererence*, 22, 137-152.

Morris, C. (1982). Natural exponential families with quadratic variance functions. *Annals of Statistics*, 10, 65-80.

Morris, C. (1983). Natural exponential families with quadratic variance functions. *Annals of Statistics*, 11, 515-529.

Prasad, N.G.N., and Rao, J.N.K. (1990). The estimation of the mean squared error of small area estimators. *Journal of the American Statistical Association*, 85, 163-171.

Sarkar, S., and Ghosh, M. (1998). Empirical Bayes estimation of local area means for NEF-QVF superpopulations. *Sankhyā*, Series B, 60, 464-487.

Tanner, M.A. (1996). *Tools for Statistical Inference*. New York: Springer.