Journal of the Korean Statistical Society (2003), **32**: 2, pp 121–150

BAYES EMPIRICAL BAYES ESTIMATION OF A PROPORTION UNDER NONIGNORABLE NONRESPONSE

JAI WON CHOI[1] AND BALGOBIN NANDRAM[2]

ABSTRACT

The National Health Interview Survey (NHIS) is one of the surveys used to assess the health status of the US population. One indicator of the nation's health is the total number of doctor visits made by the household members in the past year. There is a substantial nonresponse among the sampled households, and the main issue we address here is that the nonresponse mechanism should not be ignored because respondents and nonrespondents differ. It is standard practice to summarize the number of doctor visits by the binary variable of no doctor visit versus at least one doctor visit by a household for each of the fifty states and the District of Columbia. We consider a nonignorable nonresponse model that expresses uncertainty about ignorability through the ratio of odds of a household doctor visit among respondents to the odds of doctor visit among all households. This is a hierarchical model in which a nonignorable nonresponse model is centered on an ignorable nonresponse model. Another feature of this model is that it permits us to "borrow strength" across states as in small area estimation; this helps because some of the parameters are weakly identified. However, for simplicity we assume that the hyperparameters are fixed but unknown, and these hyperparameters are estimated by the EM algorithm; thereby making our method Bayes empirical Bayes. Our main result is that for some of the states the nonresponse mechanism can be considered nonignorable, and that 95% credible intervals of the probability of a household doctor visit and the probability that a household responds shed important light on the NHIS.

AMS 2000 subject classifications. Primary 62A15; Secondary 62D05.

Keywords. Beta-binomial model, health survey, hierarchical model, identifiability, EM algorithm, importance sampling, latent variable, selection approach, small areas.

Received September 2002; accepted March 2003.

[1]Office of Research and Methodology, National Center for Health Statistics Room 915, 6525 Belcrest Road, Hyattsville, MD 20782, U.S.A. (e-mail : jwc7@cdc.gov)

[2]Department of Mathematical Sciences, Worcester Polytechnic Institute 100 Institute Road, Worcester, MA 01609-2280, U.S.A. (e-mail : balnan@wpi.edu)

1. INTRODUCTION

Recently, nonresponse rates have been increasing in many surveys (De Heer, 1999; Groves and Couper, 1998) making the nonresponse problem more and more important. There has been much activity in estimating survey nonresponse. The main difficulty in modeling of nonresponse is in building a sensible relation between the respondents and the nonrespondents. This is especially important when there are very sparse information from the nonrespondents. While the method of ratio estimation is simple, it treats the respondents and the nonrespondents symmetrically, and therefore, inaccurate when the respondents and nonrespondents actually differ. For many surveys the units are households, and the response is binary. Thus, we propose a method to estimate the proportion of households possessing a characteristic (*e.g.*, doctor visits in the past year) using a Bayesian method which allows pooling of data across areas.

The National Health Interview Survey (NHIS) estimates the proportion of households with at least one doctor visit during the past year. The NHIS is one of the surveys used to measure the health status of the U.S. population. NHIS executes the national surveys on chronic and acute conditions, doctor visit, hospital discharge, medical care expenditure and utilization, disability and other health topics. The major part of the survey findings is included in an annual report to the public and government, and helps them formulate improved health care policies.

The main issue we consider here is how to account for the bias due to nonresponse in the NHIS. Nonresponse arises mainly from refusals, non contacts, those households with language difficulties, or households not qualified. Thus, there are differences between respondents and nonrespondents. The ratio method, used previously for the NHIS, assumes that the proportion of the characteristic for the respondents and the nonrespondents is the same. Therefore, the ratio method will be inaccurate for the situations in which respondents and nonrespondents differ.

Rubin (1987) and Little and Rubin (1987) describe two types of models which differ according to the ignorablility of nonresponse. In the ignorable model, the distribution of the variable of interest for a respondent is the same as the distribution of that variable for a nonrespondent with the same values of the covariates. In addition, the parameters in the distributions of the variable and the response must be distinct (Rubin, 1976). All other models are nonignorable. We consider a model that centers a nonignorable model on an ignorable model. In this model

an odds ratio (the odds of a household doctor visit among the responding households versus the odds of a household doctor visit among all households) is used to control the extent of nonignorability, and thereby in the Bayesian approach inducing uncertainty about ignorability. This is consistent with Draper (1995) where the overall consensus is to use a continuous model expansion whenever possible. Henceforth, we will call this family of models the expansion model because when the odds ratio is unity, the model is ignorable. Rubin (1977), Pregibon (1977), Little (1982), Nordheim (1984), Kadane (1993), Phillips (1993), and Forster and Smith (1998) use similar models that control the extent of nonignorability.

Little and Rubin (1987) and Little (1993) distinguish between two classes of models for missing data. In the selection approach the hypothetical complete data are modeled, and a model for the nonresponse mechanism is added conditional on the hypothetical data (see Heckman, 1976; Olson, 1980). In the pattern mixture approach the population is stratified into two patterns, respondents and nonrespondents, each being modeled separately and the final answer is obtained by a probabilistic mixture of these two. The selection approach is more natural for our problem.

Although the Bayesian method is appropriate for the analysis of nonignorable nonresponse problems (Little and Rubin, 1987; Rubin, 1987), the main difficulty is to model the relationship between the respondents and nonrespondents. Our objective is a simple Bayesian analysis, but a full Bayesian analysis requires Markov chain Monte Carlo (MCMC) methods which we want to avoid. Our approach here is to provide a simple algorithm in which the users need not worry about monitoring (tuning and convergence) of the algorithm. Monitoring of an MCMC algorithm can be time-consuming, needs considerable expertise, and they have to be performed for every new data set one needs to analyze. Thus, we consider the Bayes empirical Bayes approach (Deely and Lindley, 1981) to study nonignorable nonresponse. The simplicity in this approach arises, because in the posterior distributions of the parameters of interest, certain parameters (*e.g.*, hyperparameters) are assumed fixed but unknown and are estimated using the current data. In our hierarchical model we estimate the hyperparameters using the EM algorithm, and afterward we assume that these estimates are known. Indeed, this approach is useful if inference about the parameters of interest is not sensitive to the specification of the hyperparameters. However, in our case the analysis is still complicated; even modal estimates from the EM algorithm are difficult to obtain.

The NHIS data are collected from the fifty states and the District of Columbia.

For each area (a state or the District of Columbia) there are count data on the number of households, the number of responding households, and the number of household doctor visits. Like most nonresponse models many of the parameters in our expansion model are weakly identifiable. Thus, in the spirit of small area estimation, our expansion model "borrows strength" across the areas. Stasny (1991) used a hierarchical Bayes nonignorable selection model to study victimization in the National Crime Survey. A Bayes empirical Bayes approach, in which maximum likelihood estimates are substituted for the unknown hyperparameters, was used. Essentially our expansion model is an extension of her model.

Our objective is to describe the expansion model, to show how to fit it using a Bayes empirical Bayes method, and to apply it to the NHIS data on doctor visits in the past year. That is, we provide an algorithm that can be used routinely and which does not require monitoring as in MCMC studies. In Section 2 we explain the issue of nonidentifiability, and we show how the Bayesian approach provides a solution to the problem. In Section 3 we describe the Bayes empirical Bayes methodology for the expansion model. In Section 4 we present an analysis of the NHIS data, and we compare our model that pools the data with the one that uses the states individually (no pooling). Finally, Section 5 has concluding remarks.

2. ANALYSIS OF A SINGLE STATE

In this section we describe the analysis of a single state when there is nonignorable nonresponse. The purpose of this analysis is to describe the strength of the Bayesian analysis for the nonignorable nonresponse problem.

Let p denote the probability that there is at least one doctor visit in this state, π_0 the probability that a household with no doctor visit responds and π_1 the probability that a household with at least one doctor visit responds. Our basic model for nonignorable nonresponse is

$$\begin{aligned} y_j \mid p &\overset{iid}{\sim} \text{Bernoulli}(p), \\ r_j \mid \pi_0,\ y_j = 0 &\overset{iid}{\sim} \text{Bernoulli}(\pi_0), \\ r_j \mid \pi_1,\ y_j = 1 &\overset{iid}{\sim} \text{Bernoulli}(\pi_1), \end{aligned}$$

with independence over j, $j = 1, \ldots, n$, $y = \sum_{j=1}^{n} y_j$ and $r = \sum_{j=1}^{n} r_j$. Note that if $\pi_1 = \pi_0$, then the model is an ignorable nonresponse model. The parameters of interest are p, $\gamma = \pi_1/\pi_0$ and $\delta = \pi_1 p + \pi_1(1-p)$, the proportion of respondents in the entire area (population).

2.1. Nonidentifiability issue

In this section we address the issue of nonidentifiability when there is no prior information about p, π_0 and π_1.

It is easy to show that the likelihood function is

$$L(p, \pi_0, \pi_1 \mid y, r) = (\pi_1 p)^y \{\pi_0(1-p)\}^{r-y} \{1 - \pi_1 p - \pi_0(1-p)\}^{n-r}.$$

The main problem here is that in this likelihood function the numbers of households with and without doctor visits are not distinguishable among the nonrespondents. Our explanation becomes more transparent when we make the transformation $\alpha = \pi_1 p$ and $\beta = \pi_0(1-p)$. Then, we have

$$L(\alpha, \beta \mid y, r) = \alpha^y \beta^{r-y} (1 - \alpha - \beta)^{n-r}.$$

Therefore, the likelihood is a function only of two parameters, α and β. Thus, p, π_0 and π_1 are not identifiable in the likelihood function, and so they can not be estimated. Letting $\gamma = \pi_1/\pi_0$, observe that $\alpha/\beta = \gamma p/(1-p)$. Thus, if γ is known, then p is identifiable. For example, if $\gamma = 1$ (*i.e.*, $\pi_0 = \pi_1$, ignorable nonresponse model), then $p = \alpha/(\alpha + \beta)$. Thus, once γ is unknown (*i.e.*, the relation between π_0 and π_1 is unknown), p, π_0 and π_1 are all nonidentifiable. The failure of the non-Bayesian method is primarily due to the lack of information about p, π_0 and π_1. If there is some knowledge about the relation between π_0 and π_1, there will be an improvement in inference. This can be exploited through γ.

Then the main question is "How can p, π_0 and π_1 be estimated efficiently?". It turns out that an answer can be obtained, not by the traditional method, but through the Bayesian paradigm.

2.2. A Bayesian analysis

Letting z denote the number of households with at least one doctor visit among the nonrespondents. Note that z is a latent variable and is unknown to us. The entire nonresponse problem is solved once z becomes known. This is key to our analysis. Thus, starting with γ in our model, the augmented likelihood function is

$$\begin{aligned} &L(p, \gamma, \pi_0, z \mid y, r) \\ &= \binom{n-r}{z} (\gamma \pi_0 p)^y \{\pi_0(1-p)\}^{r-y} \{(1 - \gamma\pi_0)p\}^{z_i} \{(1-\pi_0)(1-p)\}^{n-r-z}, \end{aligned}$$

where $0 < \gamma < {\pi_0}^{-1}$ and $z = 0, 1, \ldots, n - r$. Suppose we consider proper non-informative prior densities for p, π_0 and γ by $p \sim U(0,1)$, $\pi_0 \sim U(0,1)$, and $\gamma \mid \pi_0 \sim U(0, {\pi_0}^{-1})$. Then the joint posterior density of the parameters z, p, π_0, γ, given y and r is

$$f(p, \gamma, \pi_0, z \mid y, r)$$
$$\propto \pi_0 \binom{n-r}{z} p^{y+z} (1-p)^{n-y-z} {\pi_0}^{r-y} (1-\pi_0)^{n-r-z} (\gamma\pi_0)^y (1-\gamma\pi_0)^z,$$

where $0 < p, \pi_0 < 1$, $0 < \gamma < {\pi_0}^{-1}$ and $z = 0, 1, \ldots, n-r$. This posterior density must be proper because the prior density is proper. It is slightly more convenient to make the transformation $p = p$, $\pi_0 = \pi_0$ and $\pi_1 = \gamma\pi_0$ for which the absolute value of the Jacobian is ${\pi_0}^{-1}$. Thus,

$$f(p, \gamma, \pi_0, z \mid y, r)$$
$$\propto \binom{n-r}{z} p^{y+z} (1-p)^{n-y-z} {\pi_0}^{r-y} (1-\pi_0)^{n-r-z} {\pi_1}^y (1-\pi_1)^z,$$

where $0 < p, \pi_0, \pi_1 < 1$ and $z = 0, 1, \ldots, n-r$. Now, letting $B(u,v) = \Gamma(u)\Gamma(v)/\Gamma(u+v)$ ($\Gamma(u)$ is the gamma function) denote the beta function, the normalization constant is

$$\sum_{z=0}^{n-r} \binom{n-r}{z} B(y+z+1, n-y-z+1) B(r-y+1, n-r-z+1)$$
$$\times B(y+1, z+1)$$

and this latter quantity is finite. That is, as indicated above, the joint posterior density of $p, \pi_0, \pi_1 \mid y, r$ is proper. Thus, p, π_0, π_1 and therefore, p, π_0,γ are all identifiable, albeit weakly.

Note that we have incorporated virtually no information through the uniform priors. Thus, it is simply because the parameters are stochastic that helps them to become identifiable. In fact, this probabilistic input in the model actually changes the structure of the model because when p, π_0, π_1 and z are integrated out, an intra-class correlation is introduced among the pairs (y_j, r_j). Another feature that is useful is that p, π_0, π_1 and z are all bounded. This is a strength of the Bayesian paradigm, and there are numerous examples in Bayesian statistics.

Inference about p, π_0 and γ can be easily obtained. For the joint posterior density of p, π_0 and π_1 is

$$f(p, \pi_0, \pi_1 \mid y, r) = \sum_{z=0}^{n-r} f(p, \pi_0, \pi_1 \mid Z = z, y, r) Pr\big(Z = z \mid y, r\big),$$

where

$$Pr(Z = z \mid y, r) = \frac{\omega_z}{\sum_{z=0}^{n-r} \omega_z}, \; z = 0, \ldots, n-r$$

and

$$\omega_z = \binom{n-r}{z} B(y+z+1, n-y-z+1) B(r-y+1, n-r-z+1) \times B(y+1, z+1).$$

Also, given z, y and r, it is clear that p, π_0 and π_1 are independent with

$$\begin{aligned} p \mid z, y, r &\sim \text{Beta}(y+z+1, n-y-z+1), \\ \pi_0 \mid z, y, r &\sim \text{Beta}(r-y+1, n-r-z+1), \\ \pi_1 \mid z, y, r &\sim \text{Beta}(y+1, z+1). \end{aligned}$$

Thus, inference about p, π_0 and π_1, and therefore δ and $\gamma = \pi_1/\pi_0$ can be made by drawing samples from $p, \pi_0, \pi_1, z \mid y, r$ using the composition method.

For the NHIS in Table 1 for each area, we present n, r and y. The last two columns show the observed proportion $\hat{p} = y/r$ of responding households with at least one doctor visit and the proportion $\hat{\delta} = r/n$ of responding households, respectively. The nine states marked with asterisks are the households with 8% or more nonrespondents. These states are Colorado, Delaware, District of Columbia, Florida, Louisiana, Maryland, New York, South Carolina and West Virginia. Hawaii and Maine reported the highest $\hat{p}$ of doctor visits with 38% for each of these states. To make inference about p, δ and γ for each individual state, we have drawn a sample of 10,000 observations using the composition method. We present 95% credible intervals for p and δ in the second and third columns of Table 2. For the moment we note that the 95% credible intervals for p and δ are wide, especially for p. In Table 3 we present summaries about the posterior densities of γ. The numerical standard error (column labeled NSE) are all very small, indicating that the computation is performing well. The posterior means (column labeled AVG) across the areas are all very similar (range: 0.92–0.98). The posterior standard deviations (column labeled STD) are very different (range: 0.032–0.172). These correspond to very wide 95% credible intervals, all containing 1. In fact, the $Pr(\gamma < 1 \mid y, r)$ lie between 0.61 and 0.73, indicating that the nonresponse mechanism is ignorable for each state. Thus, while inference can be made for a single state, such an estimate is very unreliable. If there was more information about each area, some of the 95% credible intervals may

not contain 1 and $Pr(\gamma < 1 \mid y, r)$ could be quite large (near 1). For example, consider California the 95% credible interval is $(0.85, 1.08)$ and the corresponding probability is 0.68. If more information (*e.g.*, prior) can be incorporated into the analysis, the interval will become narrower and may not contain 1. Also the corresponding probability may get closer to 1. This is discussed in detail in Section 4.

Then, the next issue is "How can we obtain more reliable estimates?". This is the key issue we address in this paper. The idea is to use a prior which permits a weighted pooling of data across the states. This is not unreasonable because one can think that there is some similarity across the states. This is the fundamental idea in small area estimation where a "borrowing of strength" across the ensemble is encouraged. We also use an expansion model which centers a nonignorable nonresponse model on an ignorable one. See Draper (1995) for a discussion of the general statistical problem, and Forster and Smith (1998) for the nonresponse problem.

3. Methodology for Small Areas

In this section we show how to pool the data from the 51 areas to improve the inference. In Section 3.1 we describe the expansion model and in Section 3.2 we describe the computations. We note that the EM algorithm is used to estimate the hyperparameters, and these are used as the true values. Then, the Metropolis algorithm is used to draw samples from the posterior distribution assuming these hyperparameters are known. This is a simplification over the full Bayesian approach in which there is a little loss.

3.1. A model for small areas

Letting ℓ be the number of areas (there are fifty one areas which are the fifty states and the District of Columbia), we assume that a sample of n_i households is taken from the i^{th} state (conveniently called area), $i = 1, \ldots, \ell$. Let the binary characteristic be

$$y_{ij} = \begin{cases} 1, & \text{if household } j \text{ in area } i \text{ visited a doctor at least once,} \\ 0, & \text{if household } j \text{ in area } i \text{ did not visit a doctor,} \end{cases}$$

and the response variable

$$r_{ij} = \begin{cases} 1, & \text{if household } j \text{ in area } i \text{ is a respondent,} \\ 0, & \text{if household } j \text{ in area } i \text{ is not a respondent,} \end{cases}$$

where $i = 1, \ldots, \ell$ and $j = 1, \ldots, n_i$, the size of the sample from the i^{th} area. Let $y_i = \sum_{j=1}^{n_i} y_{ij}$ be the number of households with at least one doctor visit, and $r_i = \sum_{j=1}^{n_i} r_{ij}$, the number of responding households over the past year. The expansion model for nonignorable nonresponse is

$$\begin{aligned} y_{ij} \mid p_i &\overset{iid}{\sim} \text{Bernoulli}(p_i), \\ r_{ij} \mid \pi_i, \gamma_i,\ y_{ij} = 1 &\overset{iid}{\sim} \text{Bernoulli}(\gamma_i \pi_i), \\ r_{ij} \mid \pi_1, \gamma_i\ y_{ij} = 0 &\overset{iid}{\sim} \text{Bernoulli}(\pi_i), \end{aligned}$$

$i = 1, \ldots, \ell,\ j = 1, \ldots, n_i$. Here γ_i is the ratio of the odds of success among respondents to the odds of success among all individuals in the i^{th} area. Observe that γ_i reflects the extent of nonignorability of the nonrespondents and, in fact, incorporate the uncertainty about ignorability into the model. If $\gamma_i = 1$, the model becomes ignorable and there is no difference between respondents and nonrespondents. One might also consider the ratio of the odds of success among respondents and the odds of success among nonrespondents (see comments in Section 4).

The parameters of interest are p_i, γ_i and δ_i where δ_i is the probability of responding in the i^{th} area and is $\delta_i = \pi_i\{\gamma_i p_i + (1 - p_i)\}$. Assuming all areas are similar, we take the parameters $(p_i, \delta_i, \gamma_i)$ to have a common distribution. This assumption is useful because it helps in the estimation for the parameters that are weakly identified by the data as described in Section 3. For p_i, we take

$$p_i \mid \mu_1, \tau_1 \overset{iid}{\sim} \text{Beta}(\mu_1 \tau_1, (1 - \mu_1)\tau_1). \tag{3.1}$$

Note that $E(p_i \mid \mu_1, \tau_1) = \mu_1$ and $\text{Var}(p_i \mid \mu_1, \tau_1) = \mu_1(1 - \mu_1)/(\tau_1 + 1)$. This reparameterization is useful because the parameters μ_1 and τ_1 are approximately orthogonal.

We wish to center the γ_i at unity (*i.e.*, center on an ignorable model). It is possible to do so by assuming that the γ_i have a common mean of unity. Thus, one can assume that $\gamma_i \mid \nu \sim iid\, \Gamma(\nu, \nu)$, $\gamma_i > 0$, where $E(\gamma_i \mid \nu) = 1$ and $\text{Var}(\gamma_i \mid \nu) = \nu^{-1}$. Thus, we can center the expansion model on an ignorable model with γ_i fluctuating about unity with a standard deviation $1/\sqrt{\nu}$ a priori.

But there is the issue that $0 < \gamma_i \pi_i < 1$. Thus, we assume that the parameters (π_i, γ_i) are jointly independent with

$$\pi_i \mid \mu_2, \tau_2 \overset{iid}{\sim} \text{Beta}(\mu_2\tau_2, (1-\mu_2)\tau_2)$$

and

$$\gamma_i \mid \nu \overset{iid}{\sim} \Gamma(\nu, \nu), \qquad 0 < \gamma_i < \pi_i^{-1}, \quad 0 < \pi_i < 1.$$

Therefore, the joint prior density for (π_i, γ_i) is given by

$$\begin{aligned} & p(\pi_i, \gamma_i \mid \mu_2, \tau_2, \nu) \qquad (3.2) \\ & = \nu\gamma_i^{\nu-1} \exp(-\nu\gamma_i) \frac{\pi_i^{\mu_2\tau_2 - 1} (1-\pi_i)^{(1-\mu_2)\tau_2 - 1}}{B(\mu_2\tau_2, (1-\mu_2)\tau_2)} I_i^{-1}(\mu_2, \tau_2, \nu), \end{aligned}$$

where letting $\phi_i = \pi_i\gamma_i$,

$$I_i(\mu_2, \tau_2, \nu) = \int_0^1 \int_0^1 \left\{ \frac{\exp(-\phi_i/\pi_i)}{\pi_i} \right\}^{\nu} f_1(\pi_i, \phi_i \mid \nu) d\pi_i d\phi_i \qquad (3.3)$$

and

$$f_1(\pi_i, \phi_i \mid \nu) = \nu\phi_i^{\nu-1} \frac{\pi_i^{\mu_2\tau_2 - 1} (1-\pi_i)^{(1-\mu_2)\tau_2 - 1}}{B(\mu_2\tau_2, (1-\mu_2)\tau_2)}, \qquad 0 < \pi_i, \phi_i < 1.$$

Thus, the joint prior distribution for (p_i, π_i, γ_i) is the product of the densities in (3.1) and (3.3).

For a full Bayesian analysis, prior distributions are needed for the hyperparameters μ_1, τ_1, μ_2, τ_2 and ν. Thus, we take

$$\mu_r \overset{iid}{\sim} \text{Beta}(1,1), \quad r = 1, 2.$$

That is, uniform proper prior densities are used for μ_1 and μ_2. We also use proper prior distributions for τ_1, τ_2 and ν. These prior distributions are similar to the uniform shrinkage proper prior distributions. Specifically, we take

$$p(\nu) = \frac{1}{(\nu+1)^2}, \quad \nu \geq 0 \quad \text{and} \quad p(\tau_r) = \frac{1}{(\tau_r+1)^2}, \quad \tau_r \geq 0, \quad r = 1, 2$$

with independence over μ_1, τ_1, μ_2, τ_2 and ν. See, for example, Albert (1988) where unity is used in the shrinkage prior distribution (*i.e.*, $\nu + 1$ instead of $\nu + a$ for some choice of a). These prior distributions discourage the posterior modal estimates of τ_1, τ_2 and ν to be on the boundary of the parameter space which will make inference difficult. We note also that Stasny (1991) essentially did not

use any prior distribution on the hyperparameters and resorted to the maximum likelihood estimators and, in fact, she did not use an expansion model.

For an empirical Bayes approach one does not need to specify the prior distributions. In this approach the hyperparameters, μ_1, τ_1, μ_2, τ_2 and ν, are simply assumed to be fixed quantities. Then, they are estimated using maximum likelihood procedures. It is slightly more elegant to provide a full Bayesian model. We have shown in Section 4 that inference about the parameters p_i, γ_i and δ_i is not sensitive to misspecifications of these hyperparameters. However, as noted earlier a full Bayesian analysis requires the implementation of Markov chain Monte Carlo methods which we avoid because we want to make our method available to practitioners. Thus, we use the Bayes empirical Bayes method in which we obtain posterior modal estimates of μ_1, τ_1, μ_2, τ_2 and ν; thereafter we treat these parameters as known quantities. We use the EM algorithm (Dempster *et al.*, 1977) to estimate these hyperparameters directly. The EM algorithm is a general approach to perform the computation of maximum likelihood estimation when the observation can be viewed as incomplete data.

Let $r_i = \sum_{j=1}^{n_i} r_{ij}$ be the number of respondents and $y_i = \sum_{j=1}^{n_i} y_{ij}$ the number of households with at least one doctor visit, and $n_i - r_i$ is the number of nonrespondents. Since the number of visits among the nonrespondents is unknown, we denote it by the latent variable z_i, and hence, the number of households with no visits among them is $n_i - r_i - z_i$.

As seen earlier for the i^{th} area, the likelihood function can be represented by a four-cell multinomial probability mass function. Then, the combined likelihood function is proportional to

$$f(\mathbf{y}, \mathbf{r}, \mathbf{z} \mid \mathbf{p}, \gamma, \pi) = \prod_{i=1}^{\ell} f(x_i, r_i, z_i \mid p_i, \gamma_i, \pi_i), \tag{3.4}$$

where

$$\begin{aligned} f(y_i, r_i, z_i \mid p_i, \gamma_i, \pi_i) = & \binom{n_i}{r_i}\binom{r_i}{y_i}\binom{n_i - r_i}{z_i}(\gamma_i \pi_i p_i)^{y_i}\left\{\pi_i(1-p_i)\right\}^{r_i - y_i} \\ & \times\left\{(1-\gamma_i\pi_i)p_i\right\}^{z_i}\left\{(1-\pi_i)(1-p_i)\right\}^{n_i - r_i - z_i}. \end{aligned}$$

Using Bayes' theorem, the joint posterior density of all the parameters $\mathbf{z}, \mathbf{p}, \boldsymbol{\pi}, \boldsymbol{\gamma}$,

$\mu_1, \tau_1, \mu_2, \tau_2, \nu$ given $\mathbf{y}$ and $\mathbf{r}$ is

$$
\begin{aligned}
&f(\mathbf{p}, \boldsymbol{\gamma}, \boldsymbol{\pi}, \mathbf{z}, \mu_1, \tau_1, \mu_2, \tau_2, \nu \mid \mathbf{y}, \mathbf{r}) \\
&\quad \propto p(\nu)\, p(\mu_1)\, p(\mu_2)\, p(\tau_1)\, p(\tau_2)\, p(\mathbf{z}) \\
&\qquad \times \prod_{i=1}^{\ell} \left\{ \binom{n_i - z_i}{z_i} \frac{p_i^{y_i + z_i + \mu_1\tau_1 - 1}(1-p_i)^{n_i - y_i - z_i + (1-\mu_1)\tau_1 - 1}}{B(\mu_1\tau_1, (1-\mu_1)\tau_1)} (\gamma_i\pi_i)^{y_i}(1-\gamma_i\pi_i)^{z_i} \right. \\
&\qquad \left. \times \frac{\pi_i^{r_i - y_i + \mu_2\tau_2 - 1}\,(1-\pi_i)^{n_i - r_i - z_i + (1-\mu_2)\tau_2 - 1}}{B(\mu_2\tau_2, (1-\mu_2)\tau_2)} \nu\gamma_i^{\nu-1} \exp(-\nu\gamma_i) I_i^{-1}(\mu_2, \tau_2, \nu) \right\},
\end{aligned}
$$

where I_i is given in (3.3). However, it is convenient to make the transformation $\phi_i = \gamma_i\pi_i$, $i = 1, \ldots, \ell$ with $\mathbf{z}$ and all other parameters untransformed. The absolute value of the Jacobian of this transformation is $\prod_{i=1}^{\ell} \pi_i^{-1}$. Therefore, the joint posterior density of all the parameters $\mathbf{z}, \mathbf{p}$, $\boldsymbol{\pi}$, $\boldsymbol{\phi}$, μ_1, τ_1, μ_2, τ_2, ν given $\mathbf{y}$ and $\mathbf{r}$ is

$$
\begin{aligned}
&f(\mathbf{p}, \boldsymbol{\phi}, \boldsymbol{\pi}, \mathbf{z}, \mu_1, \tau_1, \mu_2, \tau_2, \nu \mid \mathbf{y}, \mathbf{r}) \\
&\quad \propto p(\nu)\, p(\mu_1)\, p(\mu_2)\, p(\tau_1)\, p(\tau_2)\, p(\mathbf{z}) \\
&\qquad \times \prod_{i=1}^{\ell} \left\{ \binom{n_i - r_i}{z_i} \frac{p_i^{y_i + z_i + \mu_1\tau_1 - 1}\,(1-p_i)^{n_i - y_i - z_i + (1-\mu_1)\tau_1 - 1}}{B(\mu_1\tau_1, (1-\mu_1)\tau_1)} \right. \\
&\qquad \times \phi_i^{y_i + \nu - 1}(1-\phi_i)^{z_i} \frac{\pi_i^{r_i - y_i + \mu_2\tau_2 - 1}(1-\pi_i)^{n_i - r_i - z_i + (1-\mu_2)\tau_2 - 1}}{B(\mu_2\tau_2, (1-\mu_2)\tau_2)} \\
&\qquad \left. \times \left\{ \frac{\exp(-\phi_i/\pi_i)}{\pi_i} \right\}^{\nu} \nu I_i^{-1}(\mu_2, \tau_2, \nu) \right\} \qquad (3.5)
\end{aligned}
$$

In a full Bayesian analysis, inference proceeds after drawing a sample from $f(\mathbf{p}, \boldsymbol{\phi}, \boldsymbol{\pi}, \mathbf{z}, \tau_1, \mu_2, \tau_2, \nu \mid \mathbf{y}, \mathbf{r})$ using Markov chain Monte Carlo methods.

3.2. Computations

Because the posterior density is not accessible directly, we use a sampling based method to obtain samples from the posterior density to permit an inference. We marginalize out the parameters (p_i, π_i, ϕ_i) from the joint posterior density to obtain the posterior density $f(\mathbf{z}, \mu_1, \tau_1, \mu_2, \tau_2, \nu \mid \mathbf{y}, \mathbf{r})$. Because it is difficult to obtain samples from this posterior density, we obtain the posterior modes for the hyperparameters μ_1, τ_1, μ_2, τ_2, and ν. First, we obtain estimators in the spirit of the method of moments (see Appendix A). These estimates are to be

used to start up the EM algorithm (see Appendix B). We note that the EM algorithm converges within 10 steps and the estimates of the hyperparameters are $\mu_1 = 0.331$, $\tau_1 = 566$, $\mu_2 = 0.963$, $\tau_2 = 6099$, and $\nu = 9.018$. Then, we use these estimated values as if they are the true values of μ_1, τ_1, μ_2, τ_2 and ν, and this approach provides a Bayes empirical Bayes approach. Letting

$$\begin{aligned} & f_a(\mathbf{z}, \mu_1, \tau_1, \mu_2, \tau_2, \nu \mid \mathbf{y}, \mathbf{r}) \\ & \quad \propto p(\nu)\, p(\tau_1)\, p(\tau_2)\, p(\mathbf{z}) \\ & \qquad \times \prod_{i=1}^{\ell} \left\{ \binom{n_i - r_i}{z_i} \frac{B(y_i + z_i + \mu_1\tau_1, n_i - y_i - z_i + (1-\mu_1)\tau_1)}{B(\mu_1\tau_1, (1-\mu_1)\tau_1)} \right. \\ & \qquad \left. \times \frac{B(r_i - y_i + \mu_2\tau_2, n_i - r_i - z_i + (1-\mu_2)\tau_2)}{B(\mu_2\tau_2,\ (1-\mu_2)\tau_2)}\ \nu\ B(y_i + \nu, z_i + 1) \right\}, \end{aligned}$$

the posterior density is

$$f(\mathbf{z}, \mu_1, \tau_1, \mu_2, \tau_2, \nu \mid \mathbf{y}, \mathbf{r}) \propto f_a(\mathbf{z}, \mu_1, \tau_1, \mu_2, \tau_2, \nu \mid \mathbf{y}, \mathbf{r}) \prod_{i=1}^{\ell} \{R_{z_i}(\mu_2, \tau_2, \nu)\} \quad (3.6)$$

where

$$\begin{aligned} & R_{z_i}(\mu_2, \tau_2, \nu) \\ & \quad = I_i^{-1}(\mu_2, \tau_2, \nu) \int_0^1 \int_0^1 \left\{ \frac{\exp(-\phi_i/\pi_i)}{\pi_i} \right\}^{\nu} f_2(\pi_i, \phi_i \mid \nu, z_i, y_i, r_i) d\pi_i d\phi_i \end{aligned}$$

and

$$\begin{aligned} & f_2(\pi_i, \phi_i \mid \nu, z_i, y_i, r_i) \\ & \quad = \frac{\phi_i^{y_i+\nu-1}(1-\phi_i)^{z_i}}{B(y_i + \nu, z_i + 1)} \cdot \frac{\pi_i^{r_i - y_i + \mu_2\tau_2 - 1}\ (1-\pi_i)^{n_i - r_i - z_i + (1-\mu_2)\tau_2 - 1}}{B(r_i - y_i + \mu_2\tau_2, n_i - r_i - z_i + (1-\mu_2)\tau_2)}. \end{aligned}$$

With the known hyperparameters, we can obtain samples from the simpler joint posterior density

$$f(z_i, p_i, \pi_i, \phi_i \mid \mathbf{y}, \mathbf{r}) = g_1(p_i \mid z_i, y_i, r_i)\ g_2(\pi_i, \phi_i \mid z_i, \mathbf{y}, \mathbf{r})\ g_3(z_i \mid \mathbf{y}, \mathbf{r}).$$

The posterior density $g_1(p_i \mid z_i, y_i, r_i)$ is

$$p_i \mid y_i, r_i, z_i \overset{ind}{\sim} \text{Beta}(y_i + z_i + \mu_1\tau_1, n_i - y_i + (1-\mu_1)\tau_1) \quad (3.7)$$

from which the samples for p_i are obtained. The joint posterior density $\pi_i, \phi_i \mid y_i, r_i, z_i$ is given by

$$g_2(\pi_i, \phi_i \mid y_i, r_i, z_i) \propto \left\{ \frac{\exp(-\phi_i/\pi_i)}{\pi_i} \right\}^{\nu} g_a(\pi_i, \phi_i \mid y_i, r_i, z_i, \nu) \tag{3.8}$$

where

$$\begin{aligned} & g_a(\pi_i, \phi_i \mid z_i, y_i, r_i, \nu) \\ &= \frac{\phi_i^{y_i+\nu-1}(1-\phi_i)^{z_i}}{B(y_i+\nu, z_i+1)} \cdot \frac{\pi_i^{r_i-y_i+\mu_2\tau_2-1}\,(1-\pi_i)^{n_i-r_i-z_i+(1-\mu_2)\tau_2-1}}{B(r_i-y_i+\mu_2\tau_2, n_i-r_i-z_i+(1-\mu_2)\tau_2)}. \end{aligned}$$

The posterior probability mass functions of the z_i are

$$p(z_i \mid \mathbf{y}, \mathbf{r}) = \frac{\omega_{z_i}}{\sum_{t=0}^{n_i-r_i} \omega_t}, \quad z_i = 0, \ldots, n_i - r_i, \tag{3.9}$$

where

$$\begin{aligned} \omega_{z_i} \propto & \binom{n_i - r_i}{z_i} B(y_i + z_i + \mu_1\tau_1, n_i - y_i - z_i + (1-\mu_1)\tau_1) B(y_i + \nu, z_i + 1) \\ & \times B(r_i - y_i + \mu_2\tau_2, n_i - r_i - z_i + (1-\mu_2)\tau_2) I_{z_i}(\mu_2, \tau_2, \nu) \end{aligned}$$

and

$$\begin{aligned} I_{z_i}(\mu_2, \tau_2, \nu) = \int_0^1 \int_0^1 \Bigg\{ & \left(\frac{\exp(-\phi_i/\pi_i)}{\pi_i} \right)^{\nu} \frac{\phi_i^{y_i+\nu-1}(1-\phi_i)^{z_i}}{B(y_i+\nu, z_i+1)} \\ & \times \frac{\pi_i^{r_i-y_i+\mu_2\tau_2-1}\,(1-\pi_i)^{n_i-r_i-z_i+(1-\mu_2)\tau_2-1}}{B(r_i-y_i+\mu_2\tau_2,\ n_i-r_i-z_i+(1-\mu_2)\tau_2)} \Bigg\} d\pi_i d\phi_i. \end{aligned}$$

We first obtain z_i from equation (3.9). Then for each z_i we fill in the (p_i, π_i, ϕ_i) using equations (3.7) and (3.8). We draw 1000 values for z_i, $i = 1, \ldots, \ell$ from the equation (3.9). To draw samples from (3.8), we use Metropolis algorithm with proposal density $g_a(\pi_i, \phi_i \mid y_i, r_i, z_i, \nu)$. Assuming the chain is at the s^{th} iterate, then the jumping probability to the $(s+1)^{th}$ iterate is $A_{s,s+1} = \min\{1, (\psi(\pi_i^{(s+1)}, \phi_i^{(s+1)})/\ \psi(\pi_i^{(s)}, \phi_i^{(s)}))\}$ where

$$\psi(\pi_i, \phi_i) = \left\{ \frac{\exp(-\phi_i/\pi_i)}{\pi_i} \right\}^{\nu}.$$

Here $(\pi_i^{(s)}, \phi_i^{(s)})$ are obtained independently from

$$\phi_i^{(s)} \mid r_i, y_i, z_i, \nu \overset{ind}{\sim} \text{Beta}(y_i + \nu, z_i + 1)$$

and

$$\pi_i^{(s)} \mid r_i, y_i, z_i, \mu_2, \tau_2 \overset{ind}{\sim} \text{Beta}(r_i - y_i + \mu_2\tau_2, n_i - r_i - z_i + (1-\mu_2)\tau_2).$$

We ran the Metropolis step 100 times and we took the last one. We finally obtain a sample $(\pi_i^{(h)}, \phi_i^{(h)}, \gamma_i^{(h)})$ by taking $\gamma_i^{(h)} = \pi_i^{(h)}\phi_i^{(h)}$, $h = 1, \ldots, M$. Inference can now be made in standard way.

4. Application to the National Health Interview Survey

In this section we discuss posterior inference about the parameters of the nonignorable model for the NHIS data. We compare inference for the individual procedure and the pooled procedure. In our discussion we have summarized the posterior distribution of each of the parameters p, δ and γ for each state. To assess the computation we have also computed the numerical standard errors (obtained from the batch means method with batch length 25 for the 1,000 samples), 95% credible interval, and probability that γ is less than one. Recall that when $\gamma = 1$, there is ignorable nonresponse.

First, we consider inference for p. The fourth and the fifth columns of Table 2 show the 95% credible intervals (pooled) for p and δ by state. It is pleasant that most of the intervals for p contain the observed values (see Table 1); generally the observed values not within the 95% credible intervals are just slightly smaller. As expected, the intervals based on the pooled procedure are generally contained by those based on the individual procedure (column 2) with many of them a lot narrower. The 95% credible intervals for δ generally contain the observed values (see Table 1). Except for a few states, when the credible intervals for the pooled procedure are narrower than those for the individual procedure, the 95% credible intervals for δ are generally similar for the two procedures (columns 3 and 5). Thus, there is much gain in precision of the pooled procedure over the individual procedure for p, the parameter of greatest interest.

Second, we compare the posterior density of γ for the pooled procedure (see Table 4) with that for the individual procedure (see Table 3). First, the numerical standard errors are a lot smaller for the pooled procedure. There are some differences in the posterior mean (column labeled AVG), but the differences are generally small; one notable exception is Colorado (0.91 *vs.* 0.77). However, the posterior standard deviations (column labeled STD) are much smaller for the pooled procedure, making the intervals much narrower. Except for a few states, the $Pr(\gamma < 1 \mid y, r)$ are very close to 1, making them very different from their

counterparts in Table 3. A few of them show very small differences (*e.g.*, compare 0.65 with 0.63 for Iowa).

Third, we have also considered an alternative odds ratio to γ_i. Let γ_i^* denote the ratio of the odds of success among the respondents to odds of success among the nonrespondents for the i^{th} area. Then, it is easy to show that

$$\gamma_i^* = \frac{\gamma_i(1-\pi_i)}{1-\gamma_i\pi_i}. \tag{4.1}$$

In (4.1), γ_i^* is a strictly increasing function of γ_i on $(0, \pi_i^{-1})$ for each fixed π_i and $0 < \gamma_i^* < \infty$. That is, if $\gamma_i = 0$ then $\gamma_i^* = 0$ and if $\gamma_i \to \pi_i^{-1}$ then $\gamma_i^* \to \infty$. Thus, inference about γ_i will be similar to inference about γ_i^*. In fact, $Pr(\gamma_i < 1 \,|\, \mathbf{y}, \mathbf{r}) = Pr(\gamma_i^* < 1 \,|\, \mathbf{y}, \mathbf{r})$.

However, there are some differences between γ_i^* and γ_i. First, while γ_i is bounded above by π_i^{-1}, γ_i^* is not bounded above. Thus, as expected γ_i^* is more variable than γ_i. Second, the likelihood function for (p_i, π_i, γ_i) is simpler than the one for (p_i, π_i, γ_i^*) because the likelihood function for (p_i, π_i, γ_i^*) contains a rational function of π_i and γ_i^* as well. However, in the present model it is possible to make inference about γ_i^* by using the relation in (4.1). We note that in our applications, π_i are typically close to unity.

We compare posterior inference for γ and γ^* (Table 4 and Table 5). As explained earlier $Pr(\gamma < 1 \,|\, y, r) = Pr(\gamma^* < 1 \,|\, y, r)$ (hence the last columns in both tables are the same). In general, the posterior density for γ^* can be unstable. For example, for Idaho (a small state) compare the summaries for γ^* in Table 5 of AVG = 9.06, STD = 86.1, NSE = 11.3 and the 95% credible interval $(0.23, 38.9)$ with AVG = 1.00, STD = 0.030, NSE = 0.005 and the 95% credible interval $(0.92, 1.04)$ in Table 4. Thus, while it is sensible to make inference about γ^*, its posterior density is very unstable. Therefore, it is more beneficial to use γ to study the extent of nonignorability.

In general, it appears to be true that smaller states, with smaller number of nonrespondents, include $\gamma = 1$ in their 95% credible intervals (*i.e.*, ignorable noresponse), and that the larger states with larger number of nonrespondents, do not include $\gamma = 1$ in their 95% credible intervals (*i.e.*, nonignorable nonresponse). Some exceptions are that some smaller states, with small nonrespondents, do not include $\gamma = 1$ (Delaware with 12 nonrespondents, DC with 14 nonrespondents, WV with 22 nonrespondents). Four states are borderline, and their 95% credible intervals almost include $\gamma = 1$; they are Nebraska with 15 nonrespondents, Nevada with 14 nonrespondents, Alaska with 3 nonrespondents, and Hawaii with

13 nonrespondents. On the other hand, among the twenty states with small number of nonrespondents, some relatively large states with larger nonrespondents include $\gamma = 1$. For example, Alabama with 29 nonrespondents, Indiana with 42 nonrespondents, Minnesota with 32 nonrespondents, and Missouri with 38 nonrespondents.

It is true that if the probability that a household responds is 1, then $\gamma = 1$. Thus, by default, for a state with very few nonrespondents there must be ignorability. Some examples are Idaho, Montana, North Dakota, South Dakota, Vermont and Wyoming, states with fewer than 5 nonrespondents. For these states the 95% credible intervals for γ do contain 1, and $Pr(\gamma < 1 \,|\, y, r)$ are $0.39, 0.77, 0.90, 0.79, 0.94, 0.85$. This is consistent with expectation.

Finally, we have performed a sensitivity analysis to assess how inference is affected by the choice of the hyperparameters τ_1, τ_2 and ν. We kept μ_1 and μ_2 at the modal estimates, and set τ_1 at $100, 500, 900$, τ_2 at $3000, 6000, 9000$, and ν at $1, 9, 17$. For each of the 27 combinations, we obtained the modal estimates and we obtained samples from the joint posterior density. To summarize the results, we obtained the weighted parameters corresponding to p, γ, and δ as

$$p = \sum_{i=1}^{\ell} \tilde{n}_i p_i, \quad \gamma = \sum_{i=1}^{\ell} \tilde{n}_i \gamma_i \quad \text{and} \quad \delta = \sum_{i=1}^{\ell} \tilde{n}_i \delta_i$$

where $\tilde{n}_i = n_i / \sum_{i=1}^{\ell} n_i$, $i = 1, \ldots, \ell$. We present the results in Table 6. It is true that inference is virtually unchanged for p, γ and δ.

We have performed a full Bayesian analysis, and found that inference about p, γ and δ is similar to our empirical Bayes method. For example, for Colorado the 95% credible intervals for p, δ and γ are (0.30, 0.36), (0.87, 0.92) and (0.70, 0.84); see Table 2 and Table 4.

5. Concluding Remarks

We have presented a Bayes empirical Bayes method to estimate the proportion of doctor visits and the probability that a household responds in the NHIS. In doing so we have been able to incorporate a degree of uncertainty about the ignorability of the nonresponse mechanism. Our method assumes that the hyperparameters are fixed but unknown, and they are estimated using modal estimates from the EM algorithm. The advantage of our method over a full Bayesian approach is that it does not require monitoring of a Markov chain Monte Carlo algorithm. Thus, in our method the amount of work a practitioner needs to exert

is minimized.

Misspecifying the true values of the hyperparameters is an important issue. However, we have shown that moderate misspecification of the parameters τ_1, τ_2 and ν have little consequence on inference about p_i and δ_i as well as γ_i.

Our method is potentially useful to incorporate uncertainty about ignorability of the nonresponse mechanism for many surveys. We have shown that it is possible to decide for which states the nonresponse mechanism can be treated as ignorable. For these states it is possible to use the ratio method for nonresponse adjustment, but with some loss of precision. For the other states one must be reluctant to use the ratio method. In either case our method provides adjusted estimates with improved precision for p_i and δ_i based on the extent of nonignorability.

Our method can be extended to a full Bayesian one by modifying the algorithm of Nandram (1998). This is desirable although uncertainty about the hyperparameters might show little difference for the point estimates in our current results. Our Bayes empirical Bayes method is potentially useful to correct at least partially for nonignorable nonresponse. One other important extension is to polychotomous (more than two cells) data that are so prominent in many complex surveys.

APPENDIX

A. Method of moment estimators

First, letting $x_i = \sum_{j=1}^{n_i} r_{ij} y_{ij}$ (*i.e.*, the number of households with at least one doctor visit), we observe that

$$x_i \mid p_i, r_i \overset{ind}{\sim} \text{Binomial}(r_i, p_i) \text{ and } p_i \overset{iid}{\sim} \text{Beta}(\mu_1\tau_1, (1-\mu_1)\tau_1).$$

We obtain $E(x_i \mid r_i) = r_i\mu_1$ and

$$E(x_i^2 \mid r_i) = r_i\mu_1 + r_i(r_i - 1)\left\{\frac{\mu_1(1-\mu_1)}{\tau_1 + 1} + \mu_1^2\right\}, \quad i = 1, \ldots, \ell.$$

Then, we have

$$\hat{\mu}_1 = \frac{1}{\ell}\sum_{i=1}^{\ell} x_i.$$

Also, letting

$$t = \frac{\sum_{i=1}^{\ell} x_i^2}{\sum_{i=1}^{l} r_i^2} - \hat{\mu}_1\left(\frac{\sum_{i=1}^{\ell} r_i}{\sum_{i=1}^{\ell} r_i^2} + \hat{\mu}_i\right)$$

we define

$$\tilde{\tau}_1 = \begin{cases} \dfrac{\hat{\mu}_1(1-\hat{\mu}_1)}{t}, & \text{if } t > 0, \\ \dfrac{\hat{\mu}_1(1-\hat{\mu}_1)\sum_{i=1}^{\ell} r_i^2}{\sum_{i=1}^{\ell} x_i^2}, & \text{if } t \le 0. \end{cases}$$

Then, our estimator of τ_1 is $\hat{\tau}_1 = \max(1, \tilde{\tau}_1)$.

Second, recalling that $r_i = \sum_{j=1}^{n_i} r_{ij}$,

$$\begin{aligned} r_i \mid \pi_i, p_i, \gamma_i &\overset{ind}{\sim} \text{Binomial}(n_i, \pi_i(1-p_i) + \gamma_i \pi_i p_i), \\ p_i &\overset{iid}{\sim} \text{Beta}(\mu_1\tau_1, (1-\mu_1)\tau_1), \\ \pi_i &\overset{iid}{\sim} \text{Beta}(\mu_2\tau_2, (1-\mu_2)\tau_2), \\ \gamma_i &\overset{iid}{\sim} \text{Gamma}(\nu, \nu). \end{aligned}$$

Then, letting $\hat{\mu}_2 = \ell^{-1}\sum_{i=1}^{\ell} r_i$, approximately

$$r_i \mid \gamma_i \overset{ind}{\sim} \text{Binomial}(n_i, \hat{\mu}_2(1-\hat{\mu}_1) + \gamma_i \hat{\mu}_1 \hat{\mu}_2).$$

Thus, letting $\tilde{\gamma}_i = 1 + (r_i - \hat{\mu}_2)/\hat{\mu}_1\hat{\mu}_2$, we take

$$\hat{\gamma}_i = \begin{cases} \tilde{\gamma}_i, & \text{if } 0 < \tilde{\gamma}_i < \hat{\mu}_2^{-1}, \\ 0.5\ \hat{\mu}_2^{-1}, & \text{if } \tilde{\gamma}_i < 0 \text{ or } \tilde{\gamma}_i > \hat{\mu}_2^{-1}, \end{cases}$$

and

$$\hat{\nu} = \frac{\ell - 1}{\sum_{i=1}^{\ell} \left(\hat{\gamma}_i - \sum_{i=1}^{\ell} \hat{\gamma}_i/\ell\right)^2}.$$

Finally, letting $a_i = (1-\hat{\mu}_1) + \hat{\gamma}_i\hat{\mu}_1$, we have approximately

$$r_i \mid \pi_i \overset{ind}{\sim} \text{Binomial}(n_i, a_i\pi_i) \text{ and } \pi_i \overset{iid}{\sim} \text{Beta}(\hat{\mu}_2\tau_2, (1-\hat{\mu}_2\tau_2)).$$

Then,

$$E(r_i^2) = n_i a_i \hat{\mu}_2 + n_i(n_i - 1)a_i^2 \left\{ \frac{\hat{\mu}_2(1-\hat{\mu}_2)}{\tau_2 + 1} + \hat{\mu}_2 \right\}$$

and letting

$$t = \frac{\sum_{i=1}^{\ell} r_i^2}{\sum_{i=1}^{\ell} a_i^2 n_i^2} - \hat{\mu}_2 \left\{ \frac{\sum_{i=1}^{\ell} a_i n_i}{\sum_{i=1}^{\ell} a_i n_i^2} + \hat{\mu}_2 \right\},$$

we take

$$\tilde{\tau}_2 = \begin{cases} \dfrac{\hat{\mu}_2(1-\hat{\mu}_2)}{t}, & \text{if } t > 0, \\ \dfrac{\hat{\mu}_2(1-\hat{\mu}_2)\sum_{i=1}^{\ell} a_i^2 n_i^2}{\sum_{i=1}^{\ell} r_i^2}, & \text{if } t \le 0. \end{cases}$$

Then, our estimator of τ_2 is

$$\hat{\tau_2} = \max(1, \tilde{\tau_2}).$$

B. Expectation-maximization algorithm

The joint posterior density of μ_1, τ_1, μ_2, τ_2, ν, and $\mathbf{z}$ for given (y_i,r_i) is

$$\begin{aligned} & p(\mu_1, \tau_1, \mu_2, \tau_2, \nu, \mathbf{z} \mid \mathbf{y}, \mathbf{r}) \\ & \propto \prod_{i=1}^{\ell} \left\{ \sum_{z_i=0}^{n_i - r_i} \binom{n_i - r_i}{z_i} \frac{B(y_i + z_i + \mu_1\tau_1,\ n_i - y_i - z_i + (1-\mu_1)\tau_1)}{B(\mu_1\tau_1, (1-\mu_1)\tau_1)} \right. \\ & \quad \times \frac{B(r_i - y_i + \mu_2\tau_2, n_i - r_i - z_i + (1-\mu_2)\tau_2)}{B(\mu_2\tau_2, (1-\mu_2)\tau_2)}\ \nu B(y_i + \nu, z_i + 1) \\ & \quad \left. \times R_{z_i}(\nu, \mu_2, \tau_2) \right\} \end{aligned}$$

where

$$R_{z_i}(\nu, \mu_2, \tau_2) = \frac{\int_0^1 \int_0^1 f_1(\phi_i, \pi_i \mid y_i, r_i, z_i) d\phi_i d\pi_i}{\int_0^1 \int_0^1 f_2(\phi_i, \pi_i \mid y_i, r_i, z_i) d\phi_i d\pi_i},$$

with

$$\begin{aligned} f_1(\phi_i, \pi_i \mid y_i, r_i, z_i) = & \left\{ \frac{\exp(-\phi_i/\pi_i)}{\pi_i} \right\}^{\nu} \frac{\phi_i^{y_i + \nu - 1}\ (1-\phi_i)^{z_i}}{B(y_i + \nu,\ z_i + 1)} \\ & \times \frac{\pi_i^{r_i - y_i + \mu_2\tau_2 - 1}\ (1-\pi_i)^{n_i - r_i - z_i + (1-\mu_2)\tau_2 - 1}}{B(r_i - y_i + \mu_2\tau_2,\ n_i - r_i - z_i + (1-\mu_2)\tau_2)} \end{aligned}$$

and

$$f_2(\phi_i, \pi_i \mid y_i, r_i, z_i) = \left\{ \exp\left(-\frac{\nu\phi_i}{\pi_i} \right) \right\} \nu \phi_i^{\nu - 1}\ \frac{\pi_i^{\mu_2\tau_2 - 1}\ (1-\pi_i)^{(1-\mu_2)\tau_2 - 1}}{B(\mu_2\tau_2, (1-\mu_2)\tau_2)}.$$

Note that if z_i are given, the posterior density is separated into two parts,

$$p_1(\mu_1, \tau_1 \mid \mathbf{z}) \propto \prod_{i=1}^{\ell} \frac{B(y_i + z_i + \mu_1\tau_1,\ n_i - y_i - z_i + (1-\mu_1)\tau_1)}{B(\mu_1\tau_1, (1-\mu_1)\tau_1)}, \quad \text{(B.1)}$$

$$p_2(\mu_2, \tau_2, \nu \mid \mathbf{z}) \propto \prod_{i=1}^{\ell} A_{i1} A_{i2}, \quad \text{(B.2)}$$

where

$$A_{i1} = \frac{B(r_i - y_i + \mu_2\tau_2, n_i - r_i - z_i + (1-\mu_2)\tau_2)}{B(\mu_2\tau_2, (1-\mu_2)\tau_2)}$$

and

$$A_{i2} = \nu\ B(y_i + \nu, z_i + 1)R_{z_i}(\nu, \mu_2, \tau_2).$$

Also observe that

$$p(z_i \mid \mu_1, \tau_1, \mu_2, \tau_2, \mathbf{y}, \mathbf{r}) = \frac{\omega_{z_i}}{\sum_{t=0}^{n_i - r_i} \omega_t}, \quad z_i = 0, \ldots, n_i - r_i, \tag{B.3}$$

where

$$\omega_{z_i} \propto \binom{n_i - r_i}{z_i} B(y_i + z_i + \mu_1\tau_1, n_i - y_i - z_i + (1-\mu_1)\tau_1) B(y_i + \nu, z_i + 1)$$
$$\times B(r_i - y_i + \mu_2\tau_2, n_i - r_i - z_i + (1-\mu_2)\tau_2) I_{z_i}(\mu_2, \tau_2, \nu)$$

and

$$I_{z_i}(\mu_2, \tau_2, \nu) = \int_0^1 \int_0^1 \left\{ \left(\frac{\exp(-\phi_i/\pi_i)}{\pi_i} \right)^{\nu} \frac{\phi_i^{y_i+\nu-1}(1-\phi_i)^{z_i}}{B(y_i + \nu, z_i + 1)} \right.$$
$$\left. \times \frac{\pi_i^{r_i - y_i + \mu_2\tau_2 - 1}\ (1-\pi_i)^{n_i - r_i - z_i + (1-\mu_2)\tau_2 - 1}}{B(r_i - y_i + \mu_2\tau_2, n_i - r_i - z_i + (1-\mu_2)\tau_2)} \right\} d\pi_i d\phi_i.$$

Note that I_{z_i} does not involve the π_i and ϕ_i.

Starting with values for $\mu_1, \tau_1, \mu_2, \tau_2, \nu$ obtained from the EM algorithm, we compute the expected value of z_i from (B.3). The integral $I_{z_i}(\mu_2, \tau_2, \nu)$ is obtained using importance sample with importance function

$$\frac{\phi_i^{y_i+\nu-1}(1-\phi_i)^{z_i}}{B(y_i + \nu,\ z_i + 1)} \cdot \frac{\pi_i^{r_i - y_i + \mu_2\tau_2 - 1}\ (1-\pi_i)^{n_i - r_i - z_i + (1-\mu_2)\tau_2 - 1}}{B(r_i - y_i + \mu_2\tau_2,\ n_i - r_i - z_i + (1-\mu_2)\tau_2)},$$

where

$$\phi_i \mid y_i, r_i, z_i \overset{ind}{\sim} \text{Beta}(y_i + \nu, z_i + 1)$$

and independently

$$\pi_i \mid y_i, r_i, z_i \overset{ind}{\sim} \text{Beta}(r_i - y_i + \mu_2\tau_2, n_i - r_i - z_i + (1-\mu_2)\tau_2).$$

The EM algorithm is constructed as follows. First, the expected values of z_i are easily obtained from the probability mass function in (B.3). These are then substituted into (B.1) and (B.2). Second, (B.1) is maximized for μ_1 and τ_1 and (B.2) is maximized for μ_2, τ_2 and ν. We iterate this procedure until convergence which is rapid. The expectation step is the time consuming because the Monte Carlo integration required, but maximization step is fast. The convergence is reached within 10 iterations.

TABLE 1 *Observed proportion $\hat{p}$ of households with doctor visits and the proportion $\hat{\delta}$ of respondents for the 1995 NHIS data by state*

State	y	$r-y$	$n-r$	$\hat{p}$	$\hat{\delta}$
Alabama	222	427	29	0.34	0.96
Alaska	15	29	3	0.34	0.94
Arizona	175	421	36	0.29	0.94
Arkansas	90	252	12	0.26	0.97
California	1561	3415	289	0.31	0.95
Colorado*	144	385	62	0.27	0.90
Connecticut	170	386	31	0.31	0.95
Delaware*	37	63	12	0.37	0.89
DC*	31	66	14	0.32	0.87
Florida*	706	1542	219	0.31	0.91
Georgia	294	693	62	0.30	0.94
Hawaii	76	122	13	0.38	0.94
Idaho	44	106	1	0.29	0.99
Illinois	478	1220	131	0.28	0.93
Indiana	281	617	42	0.31	0.96
Iowa	131	234	13	0.36	0.97
Kansas	111	264	20	0.30	0.95
Kentucky	198	382	38	0.34	0.94
Louisiana*	186	389	54	0.32	0.91
Maine	63	103	11	0.38	0.94
Maryland*	223	446	69	0.33	0.91
Massachusetts	268	642	53	0.29	0.94
Michigan	473	907	92	0.34	0.94
Minnesota	223	439	32	0.34	0.95
Mississippi	99	241	14	0.29	0.96
Missouri	260	594	38	0.30	0.96
Montana	40	104	4	0.28	0.97
Nebraska	78	162	15	0.33	0.94
Nevada	60	113	14	0.35	0.93
New Hampshire	49	109	8	0.31	0.95
New Jersey	383	834	92	0.31	0.93
New Mexico	112	282	16	0.28	0.96
New York*	860	1962	278	0.30	0.91
North Carolina	310	707	72	0.30	0.93
North Dakota	29	61	4	0.32	0.96
Ohio	520	1019	75	0.34	0.95
Oklahoma	169	369	35	0.31	0.94
Oregon	134	362	12	0.27	0.98
Pennsylvania	567	1111	117	0.34	0.93
Rhode Island	43	109	6	0.28	0.96
South Carolina*	153	311	43	0.33	0.92
South Dakota	32	78	3	0.29	0.97

(*continued*)

TABLE 1 *Observed proportion $\hat{p}$ of households with doctor visits and the proportion $\hat{\delta}$ of respondents for the 1995 NHIS data by state (continued)*

State	y	$r-y$	$n-r$	$\hat{p}$	$\hat{\delta}$
Tennessee	257	535	42	0.32	0.95
Texas	949	2282	187	0.29	0.95
Utah	58	163	11	0.26	0.95
Vermont	31	66	5	0.32	0.95
Virginia	312	621	63	0.33	0.94
Washington	259	517	47	0.33	0.94
West Virginia*	68	143	22	0.32	0.91
Wisconsin	244	499	22	0.33	0.97
Wyoming	21	38	2	0.36	0.97

NOTE : **States with 8% or more nonrespondents. Here, n is the sample size, r is the number of responding households, and y is the number of households with at least one doctor visit among the respondents.*

TABLE 2 *95% credible intervals for p and δ for the individual and the pooled procedures by state*

	Individual		*Pooled*	
State	p	δ	p	δ
Alabama	(0.308, 0.394)	(0.937, 0.969)	(0.313, 0.367)	(0.939, 0.969)
Alaska	(0.223, 0.500)	(0.808, 0.967)	(0.301, 0.373)	(0.858, 0.965)
Arizona	(0.258, 0.353)	(0.920, 0.957)	(0.296, 0.349)	(0.919, 0.955)
Arkansas	(0.224, 0.324)	(0.939, 0.979)	(0.275, 0.339)	(0.931, 0.973)
California	(0.293, 0.354)	(0.938, 0.951)	(0.317, 0.343)	(0.938, 0.950)
Colorado*	(0.231, 0.361)	(0.866, 0.916)	(0.302, 0.359)	(0.866, 0.913)
Connecticut	(0.269, 0.365)	(0.924, 0.961)	(0.300, 0.354)	(0.923, 0.959)
Delaware*	(0.280, 0.495)	(0.813, 0.932)	(0.310, 0.383)	(0.853, 0.941)
DC*	(0.233, 0.460)	(0.793, 0.918)	(0.306, 0.379)	(0.828, 0.923)
Florida*	(0.283, 0.379)	(0.899, 0.921)	(0.330, 0.366)	(0.900, 0.921)
Georgia	(0.268, 0.353)	(0.924, 0.952)	(0.296, 0.346)	(0.923, 0.951)
Hawaii	(0.316, 0.467)	(0.892, 0.960)	(0.315, 0.387)	(0.904, 0.962)
Idaho	(0.227, 0.372)	(0.954, 0.996)	(0.290, 0.358)	(0.941, 0.986)
Illinois	(0.255, 0.341)	(0.915, 0.939)	(0.298, 0.336)	(0.914, 0.937)
Indiana	(0.283, 0.361)	(0.939, 0.966)	(0.299, 0.349)	(0.938, 0.965)
Iowa	(0.314, 0.417)	(0.939, 0.978)	(0.314, 0.372)	(0.939, 0.975)
Kansas	(0.253, 0.362)	(0.921, 0.965)	(0.294, 0.353)	(0.919, 0.962)
Kentucky	(0.301, 0.403)	(0.915, 0.954)	(0.318, 0.375)	(0.919, 0.955)
Louisiana*	(0.281, 0.399)	(0.888, 0.932)	(0.318, 0.374)	(0.891, 0.933)

(*continued*)

TABLE 2 *95% credible intervals for p and δ for the individual and the pooled procedures by state (continued)*

	Individual		*Pooled*	
State	p	δ	p	δ
Maine	(0.307, 0.470)	(0.885, 0.961)	(0.313, 0.383)	(0.899, 0.965)
Maryland*	(0.289, 0.409)	(0.882, 0.924)	(0.328, 0.384)	(0.888, 0.927)
Massachusetts	(0.265, 0.348)	(0.928, 0.957)	(0.295, 0.344)	(0.926, 0.954)
Michigan	(0.312, 0.394)	(0.923, 0.949)	(0.331, 0.375)	(0.925, 0.948)
Minnesota	(0.301, 0.389)	(0.934, 0.966)	(0.311, 0.367)	(0.935, 0.966)
Mississippi	(0.249, 0.354)	(0.932, 0.974)	(0.291, 0.349)	(0.926, 0.970)
Missouri	(0.274, 0.352)	(0.941, 0.968)	(0.296, 0.347)	(0.940, 0.966)
Montana	(0.216, 0.363)	(0.924, 0.985)	(0.286, 0.359)	(0.919, 0.974)
Nebraska	(0.270, 0.404)	(0.900, 0.961)	(0.303, 0.368)	(0.903, 0.961)
Nevada	(0.279, 0.443)	(0.873, 0.951)	(0.310, 0.378)	(0.888, 0.954)
New Hampshire	(0.249, 0.399)	(0.902, 0.972)	(0.296, 0.366)	(0.904, 0.969)
New Jersey	(0.283, 0.372)	(0.914, 0.942)	(0.316, 0.358)	(0.915, 0.941)
New Mexico	(0.245, 0.345)	(0.935, 0.974)	(0.286, 0.346)	(0.929, 0.970)
New York*	(0.275, 0.370)	(0.899, 0.920)	(0.326, 0.359)	(0.900, 0.919)
North Carolina	(0.273, 0.363)	(0.917, 0.947)	(0.308, 0.354)	(0.916, 0.945)
North Dakota	(0.239, 0.434)	(0.882, 0.977)	(0.293, 0.367)	(0.892, 0.973)
Ohio	(0.312, 0.380)	(0.942, 0.962)	(0.324, 0.364)	(0.942, 0.963)
Oklahoma	(0.276, 0.378)	(0.915, 0.954)	(0.305, 0.361)	(0.915, 0.953)
Oregon	(0.236, 0.318)	(0.957, 0.985)	(0.277, 0.331)	(0.952, 0.976)
Pennsylvania	(0.308, 0.390)	(0.922, 0.945)	(0.333, 0.372)	(0.923, 0.946)
Rhode Island	(0.222, 0.370)	(0.913, 0.979)	(0.288, 0.358)	(0.904, 0.972)
South Carolina*	(0.283, 0.408)	(0.886, 0.934)	(0.320, 0.377)	(0.890, 0.935)
South Dakota	(0.217, 0.390)	(0.913, 0.985)	(0.290, 0.362)	(0.912, 0.975)
Tennessee	(0.291, 0.376)	(0.931, 0.962)	(0.309, 0.363)	(0.931, 0.961)
Texas	(0.273, 0.338)	(0.937, 0.952)	(0.297, 0.328)	(0.936, 0.951)
Utah	(0.213, 0.342)	(0.912, 0.970)	(0.285, 0.351)	(0.907, 0.965)
Vermont	(0.239, 0.429)	(0.879, 0.973)	(0.298, 0.371)	(0.887, 0.970)
Virginia	(0.301, 0.389)	(0.919, 0.950)	(0.321, 0.370)	(0.921, 0.950)
Washington	(0.298, 0.390)	(0.923, 0.956)	(0.316, 0.367)	(0.925, 0.956)
West Virginia*	(0.260, 0.421)	(0.858, 0.934)	(0.312, 0.375)	(0.871, 0.936)
Wisconsin	(0.297, 0.371)	(0.955, 0.980)	(0.304, 0.355)	(0.955, 0.976)
Wyoming	(0.245, 0.489)	(0.871, 0.983)	(0.300, 0.372)	(0.888, 0.975)

NOTE : *p is the proportion of households with at least one visit and δ the proportion of respon dents.*

TABLE 3 *Summaries for posterior density of γ for the individual procedure by state*

State	*AVG*	*STD*	*NSE*	*Interval*	*Probability*
Alabama	0.971	0.058	0.009	(0.871, 1.071)	0.652
Alaska	0.949	0.130	0.021	(0.687, 1.195)	0.638
Arizona	0.950	0.079	0.010	(0.815, 1.084)	0.693
Arkansas	0.957	0.057	0.008	(0.845, 1.052)	0.732
California	0.960	0.070	0.010	(0.847, 1.078)	0.678
Colorado*	0.909	0.138	0.020	(0.687, 1.155)	0.709
Connecticut	0.956	0.074	0.010	(0.829, 1.082)	0.682
Delaware*	0.953	0.147	0.022	(0.695, 1.228)	0.615
DC*	0.922	0.172	0.025	(0.627, 1.244)	0.656
Florida*	0.940	0.110	0.015	(0.767, 1.132)	0.668
Georgia	0.952	0.079	0.012	(0.819, 1.086)	0.680
Hawaii	0.972	0.085	0.013	(0.816, 1.128)	0.610
Idaho	0.983	0.032	0.004	(0.907, 1.034)	0.688
Illinois	0.936	0.095	0.013	(0.784, 1.100)	0.705
Indiana	0.965	0.061	0.009	(0.860, 1.068)	0.674
Iowa	0.979	0.050	0.007	(0.885, 1.065)	0.632
Kansas	0.954	0.074	0.010	(0.821, 1.080)	0.688
Kentucky	0.961	0.080	0.011	(0.825, 1.100)	0.652
Louisiana*	0.943	0.109	0.016	(0.765, 1.135)	0.666
Maine	0.971	0.089	0.012	(0.808, 1.133)	0.607
Maryland*	0.944	0.117	0.016	(0.755, 1.151)	0.653
Massachusetts	0.952	0.074	0.010	(0.827, 1.079)	0.697
Michigan	0.962	0.078	0.011	(0.835, 1.096)	0.649
Minnesota	0.969	0.061	0.009	(0.863, 1.075)	0.656
Mississippi	0.960	0.062	0.009	(0.844, 1.065)	0.701
Missouri	0.963	0.059	0.008	(0.861, 1.064)	0.688
Montana	0.961	0.060	0.009	(0.834, 1.060)	0.709
Nebraska	0.956	0.084	0.012	(0.804, 1.105)	0.665
Nevada	0.954	0.104	0.015	(0.767, 1.139)	0.664
New Hampshire	0.955	0.079	0.011	(0.799, 1.093)	0.679
New Jersey	0.950	0.090	0.012	(0.804, 1.105)	0.667
New Mexico	0.960	0.061	0.009	(0.847, 1.062)	0.705
New York*	0.934	0.111	0.017	(0.760, 1.131)	0.680
North Carolina	0.948	0.086	0.013	(0.806, 1.097)	0.683
North Dakota	0.959	0.084	0.013	(0.789, 1.108)	0.660
Ohio	0.969	0.059	0.009	(0.871, 1.070)	0.660
Oklahoma	0.953	0.082	0.013	(0.812, 1.096)	0.686
Oregon	0.971	0.040	0.006	(0.894, 1.038)	0.716
Pennsylvania	0.961	0.081	0.011	(0.829, 1.101)	0.650
Rhode Island	0.954	0.071	0.010	(0.813, 1.076)	0.707
South Carolina*	0.946	0.109	0.015	(0.765, 1.136)	0.657
South Dakota	0.962	0.064	0.009	(0.824, 1.071)	0.697

(*continued*)

TABLE 3 *Summaries for posterior density of γ for the individual procedure by state (continued)*

State	*AVG*	*STD*	*NSE*	*Interval*	*Probability*
Tennessee	0.963	0.067	0.010	(0.849, 1.077)	0.667
Texas	0.953	0.072	0.011	(0.837, 1.076)	0.700
Utah	0.943	0.079	0.011	(0.791, 1.076)	0.724
Vermont	0.955	0.090	0.014	(0.773, 1.115)	0.668
Virginia	0.958	0.081	0.010	(0.825, 1.098)	0.660
Washington	0.963	0.074	0.010	(0.838, 1.091)	0.654
West Virginia*	0.938	0.125	0.017	(0.725, 1.163)	0.664
Wisconsin	0.978	0.041	0.005	(0.903, 1.047)	0.665
Wyoming	0.967	0.086	0.013	(0.790, 1.125)	0.639

TABLE 4 *Summaries for posterior density of γ for the pooled procedure by state*

State	*AVG*	*STD*	*NSE*	*Interval*	*Probability*
Alabama	0.969	0.023	0.003	(0.923, 1.012)	0.920
Alaska	0.841	0.087	0.008	(0.673, 0.998)	0.977
Arizona	0.911	0.030	0.004	(0.852, 0.969)	0.999
Arkansas	0.965	0.036	0.004	(0.885, 1.025)	0.833
California	0.933	0.012	0.001	(0.910, 0.957)	1.000
Colorado*	0.768	0.035	0.006	(0.697, 0.831)	1.000
Connecticut	0.928	0.030	0.005	(0.865, 0.984)	0.992
Delaware*	0.804	0.065	0.010	(0.674, 0.929)	1.000
DC*	0.736	0.071	0.012	(0.592, 0.873)	1.000
Florida*	0.839	0.017	0.002	(0.805, 0.872)	1.000
Georgia	0.913	0.024	0.004	(0.865, 0.958)	1.000
Hawaii	0.916	0.043	0.005	(0.822, 0.992)	0.984
Idaho	1.001	0.031	0.005	(0.919, 1.036)	0.388
Illinois	0.868	0.020	0.003	(0.830, 0.907)	1.000
Indiana	0.959	0.022	0.003	(0.916, 1.003)	0.967
Iowa	0.988	0.027	0.003	(0.922, 1.031)	0.653
Kansas	0.926	0.035	0.005	(0.847, 0.989)	0.986
Kentucky	0.917	0.027	0.003	(0.861, 0.968)	0.999
Louisiana*	0.847	0.029	0.004	(0.782, 0.903)	1.000
Maine	0.915	0.049	0.006	(0.806, 1.000)	0.974
Maryland*	0.834	0.028	0.004	(0.776, 0.884)	1.000
Massachusetts	0.922	0.024	0.003	(0.872, 0.969)	1.000
Michigan	0.918	0.018	0.002	(0.882, 0.952)	1.000
Minnesota	0.959	0.025	0.004	(0.907, 1.006)	0.951
Mississippi	0.955	0.037	0.005	(0.881, 1.018)	0.896
Missouri	0.962	0.023	0.004	(0.912, 1.004)	0.953
Montana	0.962	0.046	0.006	(0.856, 1.031)	0.773

(*continued*)

TABLE 4 *Summaries for posterior density of γ for the pooled procedure by state (continued)*

State	*AVG*	*STD*	*NSE*	*Interval*	*Probability*
Nebraska	0.909	0.044	0.006	(0.811, 0.987)	0.988
Nevada	0.872	0.051	0.009	(0.764, 0.970)	0.998
New Hampshire	0.919	0.053	0.007	(0.810, 1.015)	0.943
New Jersey	0.889	0.021	0.003	(0.850, 0.929)	1.000
New Mexico	0.958	0.033	0.005	(0.885, 1.021)	0.897
New York*	0.832	0.015	0.002	(0.802, 0.862)	1.000
North Carolina	0.894	0.022	0.003	(0.850, 0.937)	1.000
North Dakota	0.918	0.066	0.009	(0.773, 1.021)	0.901
Ohio	0.962	0.017	0.002	(0.926, 0.995)	0.988
Oklahoma	0.907	0.030	0.004	(0.848, 0.961)	0.997
Oregon	1.001	0.022	0.003	(0.953, 1.035)	0.446
Pennsylvania	0.910	0.018	0.002	(0.875, 0.946)	1.000
Rhode Island	0.932	0.057	0.007	(0.805, 1.025)	0.886
South Carolina*	0.849	0.033	0.004	(0.781, 0.910)	1.000
South Dakota	0.952	0.054	0.008	(0.829, 1.031)	0.787
Tennessee	0.943	0.024	0.003	(0.895, 0.987)	0.996
Texas	0.928	0.015	0.002	(0.899, 0.957)	1.000
Utah	0.908	0.049	0.008	(0.806, 1.000)	0.975
Vermont	0.902	0.067	0.011	(0.764, 1.016)	0.940
Virginia	0.912	0.022	0.003	(0.867, 0.958)	1.000
Washington	0.928	0.024	0.003	(0.880, 0.972)	1.000
West Virginia*	0.816	0.051	0.007	(0.715, 0.912)	1.000
Wisconsin	1.006	0.018	0.002	(0.964, 1.034)	0.320
Wyoming	0.928	0.070	0.012	(0.771, 1.029)	0.851

TABLE 5 *Summaries for posterior density of γ^* for the pooled procedure by state*

State	*AVG*	*STD*	*NSE*	*Interval*	*Probability*
Alabama	0.612	0.371	0.040	(0.286, 1.459)	0.920
Alaska	0.339	2.728	0.384	(0.065, 0.962)	0.977
Arizona	0.284	0.101	0.010	(0.160, 0.539)	0.999
Arkansas	0.860	2.031	0.272	(0.204, 3.337)	0.833
California	0.330	0.056	0.007	(0.236, 0.459)	1.000
Colorado*	0.106	0.021	0.004	(0.070, 0.153)	1.000
Connecticut	0.342	0.145	0.018	(0.175, 0.684)	0.992
Delaware*	0.142	0.068	0.009	(0.065, 0.318)	1.000
DC*	0.097	0.038	0.007	(0.045, 0.196)	1.000
Florida* 10	0.155	0.022	0.004	(0.117, 0.204)	1.000
Georgia	0.280	0.077	0.013	(0.171, 0.458)	1.000
Hawaii	0.346	0.353	0.059	(0.131, 0.817)	0.984

(*continued*)

TABLE 5 *Summaries for posterior density of γ^* for the pooled procedure by state (continued)*

State	*AVG*	*STD*	*NSE*	*Interval*	*Probability*
Idaho	9.060	86.112	11.292	(0.232, 38.905)	0.388
Illinois	0.188	0.034	0.005	(0.134, 0.263)	1.000
Indiana	0.524	0.595	0.081	(0.262, 1.076)	0.967
Iowa	1.470	4.417	0.797	(0.290, 5.993)	0.653
Kansas	0.359	0.388	0.059	(0.161, 0.766)	0.986
Kentucky	0.298	0.100	0.012	(0.171, 0.521)	0.999
Louisiana*	0.167	0.037	0.005	(0.110, 0.252)	1.000
Maine 10	0.404	1.176	0.174	(0.127, 1.008)	0.974
Maryland*	0.153	0.030	0.004	(0.100, 0.217)	1.000
Massachusetts	0.307	0.093	0.012	(0.180, 0.535)	1.000
Michigan	0.287	0.057	0.008	(0.197, 0.420)	1.000
Minnesota	0.533	0.475	0.075	(0.246, 1.201)	0.951
Mississippi	0.762	2.841	0.386	(0.193, 2.207)	0.896
Missouri	0.594	2.147	0.312	(0.258, 1.134)	0.953
Montana	1.940	13.418	1.994	(0.170, 5.758)	0.773
Nebraska	0.310	0.227	0.030	(0.127, 0.742)	0.988
Nevada	0.224	0.122	0.022	(0.099, 0.533)	0.998
New Hampshire 10	0.556	3.616	0.487	(0.127, 1.702)	0.943
New Jersey	0.224	0.045	0.007	(0.154, 0.327)	1.000
New Mexico	0.651	1.134	0.145	(0.210, 2.321)	0.897
New York*	0.148	0.021	0.003	(0.112, 0.192)	1.000
North Carolina	0.233	0.050	0.007	(0.154, 0.344)	1.000
North Dakota	0.554	1.067	0.133	(0.102, 2.535)	0.901
Ohio	0.499	0.155	0.020	(0.296, 0.883)	0.988
Oklahoma	0.274	0.102	0.010	(0.156, 0.487)	0.997
Oregon	3.625	17.381	2.262	(0.411, 18.676)	0.446
Pennsylvania	0.269	0.053	0.006	(0.186, 0.402)	1.000
Rhode Island 10	1.063	6.146	0.788	(0.122, 3.152)	0.886
South Carolina*	0.172	0.042	0.005	(0.107, 0.275)	1.000
South Dakota	1.571	13.754	1.934	(0.139, 7.220)	0.787
Tennessee	0.397	0.166	0.019	(0.223, 0.749)	0.996
Texas	0.315	0.062	0.008	(0.217, 0.465)	1.000
Utah	0.336	0.357	0.040	(0.119, 0.998)	0.975
Vermont	0.444	0.896	0.126	(0.099, 1.935)	0.940
Virginia	0.276	0.068	0.010	(0.177, 0.464)	1.000
Washington	0.329	0.097	0.011	(0.195, 0.567)	1.000
West Virginia*	0.145	0.052	0.008	(0.077, 0.265)	1.000
Wisconsin 10	2.988	10.954	1.456	(0.480, 14.614)	0.320
Wyoming	1.304	8.525	1.181	(0.103, 5.597)	0.851

TABLE 6 *Sensitivity of the 95% credible intervals for p, γ, and δ by choice of τ_1, τ_2, and ν*

τ_1	τ_2	ν	p	γ	δ
100	3000	1	(0.314, 0.330)	(0.929, 0.981)	(0.940, 0.949)
		9	(0.314, 0.330)	(0.929, 0.982)	(0.940, 0.950)
		17	(0.314, 0.330)	(0.929, 0.979)	(0.940, 0.949)
	6000	1	(0.315, 0.331)	(0.927, 0.976)	(0.939, 0.951)
		9	(0.315, 0.331)	(0.927, 0.977)	(0.939, 0.951)
		17	(0.314, 0.331)	(0.928, 0.978)	(0.939, 0.951)
	9000	1	(0.314, 0.331)	(0.928, 0.976)	(0.939, 0.952)
		9	(0.314, 0.331)	(0.929, 0.974)	(0.939, 0.952)
		17	(0.314, 0.331)	(0.929, 0.975)	(0.940, 0.952)
500	3000	1	(0.318, 0.332)	(0.929, 0.980)	(0.939, 0.949)
		9	(0.318, 0.331)	(0.929, 0.981)	(0.940, 0.950)
		17	(0.318, 0.332)	(0.929, 0.980)	(0.940, 0.950)
	6000	1	(0.318, 0.332)	(0.927, 0.975)	(0.940, 0.951)
		9	(0.318, 0.332)	(0.928, 0.975)	(0.940, 0.951)
		17	(0.318, 0.332)	(0.928, 0.975)	(0.939, 0.951)
	9000	1	(0.318, 0.332)	(0.928, 0.974)	(0.939, 0.952)
		9	(0.318, 0.332)	(0.928, 0.974)	(0.939, 0.951)
		17	(0.318, 0.332)	(0.928, 0.975)	(0.940, 0.952)
900	3000	1	(0.319, 0.332)	(0.928, 0.982)	(0.940, 0.950)
		9	(0.320, 0.332)	(0.928, 0.981)	(0.940, 0.950)
		17	(0.319, 0.332)	(0.929, 0.980)	(0.940, 0.950)
	6000	1	(0.319, 0.332)	(0.927, 0.976)	(0.939, 0.951)
		9	(0.319, 0.332)	(0.929, 0.975)	(0.940, 0.951)
		17	(0.319, 0.332)	(0.927, 0.976)	(0.940, 0.951)
	9000	1	(0.320, 0.332)	(0.927, 0.974)	(0.940, 0.951)
		9	(0.319, 0.332)	(0.927, 0.975)	(0.940, 0.952)
		17	(0.319, 0.332)	(0.928, 0.976)	(0.940, 0.952)

NOTE : *Inference about p, γ and δ is not sensitive to the misspecifications of the hyperparameters.*

REFERENCES

ALBERT, J. (1988). "Bayesian estimation of Poisson means using hierarchical log-linear model", In *Bayesian Statistics 3rd* (J. M. Bernardo, M. H. DeGroat, D. V. Lindley and A. F. M. Smith eds.), 519–531, Oxford University Press, New York.

DEELY, J. J. AND LINDLEY, D. V. (1981). "Bayes empirical Bayes", *Journal of the American Statistical Association*, **76**, 833–841.

DE HEER, W. (1999). "International response trends : Results of an international survey", *Journal of Official Statistics*, **15**, 129–142.

DEMPSTER, A. P., LAIRD, N. M. AND RUBIN, D. B. (1977). "Maximum likelihood from incomplete data via the EM algorithm", *Journal of the Royal Statistical Society*, **B39**, 1–22.

DRAPER, D. (1995). "Assessment and propagation of model uncertainty (with discussion)", *Journal of the Royal Statistical Society*, **B57**, 45–97.

FORSTER, J. J. AND SMITH, P. W. F. (1998). "Model-based inference for categorical survey data subject to non-ignorable non-response", *Journal of the Royal Statistical Society*, **B60**, 57–70.

GROVES, R. M. AND COUPER, M. P. (1998). *Nonresponse in Household Interview Surveys*, Wiley, New York.

HECKMAN, J. (1976). " The common structure of statistical models of truncation, sample selection and limited dependent variables and a simple estimator for such models", *Annals of Economic and Social Measurement*, **5**, 475–492.

KADANE, J. B. (1993). "Subjective Bayesian analysis for surveys with missing data", *Statistician*, **42**, 415–426.

LITTLE, R. J. A. (1982). "Models for nonresponse in sample surveys", *Journal of the American Statistical Association*, **77**, 237–250.

LITTLE, R. J. A. (1993). "Pattern-mixture models for multivariate incomplete data", *Journal of the American Statistical Association*, **88**, 125–134.

LITTLE, R. J. A. AND RUBIN, D. B. (1987). *Statistical Analysis with Missing Data*, Wiley, New York.

NANDRAM, B. (1998). "A Bayesian analysis of the three-stage hierarchical multinomial model", *Journal of Statistical Computation and Simulation*, **61**, 97–126.

NORDHEIM, E. V. (1984). "Inference from nonrandomly missing categorical data : An example from a genetic study on Turner's syndrome", *Journal of the American Statistical Association*, **79**, 772–780.

OLSON, R. L. (1980). "A least square correction for selectivity bias", *Econometrica*, **48**, 1815–1820.

PHILLIPS, M. J. (1993). "Contingency tables with missing data", *Statistician*, **42**, 415–426.

PREGIBON, D. (1977). "Typical survey data : Estimation and imputation", *Survey Methodology*, **2**, 70–102.

RUBIN, D. B. (1976). "Inference and missing data", *Biometrika*, **63**, 581–590.

RUBIN, D. B. (1977). "Formalizing subjective notions about the effect of nonrespondents in sample surveys", *Journal of the American Statistical Association*, **72**, 538–543.

RUBIN, D. B. (1987). *Multiple Imputation for Nonresponse in Surveys*, Wiley, New York.

STASNY, E. A. (1991). "Hierarchical models for the probabilities of a survey classification and nonresponse : An example from the national crime survey", *Journal of the American Statistical Association*, **86**, 296–303.