A Hierarchical Bayesian Nonresponse Model for Binary Data with Uncertainty about Ignorability

Balgobin Nandram, Worcester Polytechnic Institute (WPI)
Jai Won Choi, National Center for Health Statistics
B. Nandram, Dept. of Math. Sciences, WPI, 100 Institute Rd., Worcester, MA 01609

Key Words: Complex survey, Model uncertainty, Proportion, Selection model.

Abstract: First, we consider two Bayesian hierarchical models for binary nonresponse data which are clustered within a number of areas. While the first model assumes that the nonresponse mechanism is ignorable, the second assumes it to be nonignorable. Then, we introduce our best model through a continuous model expansion on an odds ratio (odds of success among respondents versus odds of success among all individuals) for each area. When the odds ratio is one, we have the ignorable model, otherwise the model is nonignorable. One important feature is that uncertainty about ignorability is incorporated by "centering" on the ignorable model. Thus, posterior inference about the odds ratio permits us to make a decision about ignorability. Our methodology is used to analyze data from the National Health Interview Survey (NHIS), a household survey, in which the areas are states. The complexity of the posterior distributions of the parameters forces us to implement the methodology using Markov chain Monte Carlo methods. We found that there are differences among the three models for estimating the proportion of households with a characteristic (doctor visit in the NHIS) and the response probability. The expansion model provides evidence that nonresponse for most of the areas is informative.

1. INTRODUCTION

We consider the problem of Bayesian modeling of nonignorable nonresponse (Rubin 1987) for binary data from a number of similar areas some of which may be small relative to the others, and we attempt to solve this problem using multi-stage hierarchical Bayesian modeling. Stasny (1991) described a Bayes empirical Bayes approach for two selection models and Nandram and Choi (2000) discussed how to use Markov chain Monte Carlo methods to provide a full Bayesian analysis.

We wish to make two important contributions. First, for these two models we address the issue of whether to use a discrete model expansion or a continuous model expansion for expressing uncertainty about ignorability in nonresponse. Second, the key point in this paper is to show how to incorporate uncertainty about ignorability using a single model in which the degree of ignorability varies from one area to the other.

Forster and Smith (1998) use a pattern mixture specification for multinomial-Dirichlet graphical models. One drawback of their method is that the prior density of the parameter which controls the extent of ignorability is the same as the posterior density of that parameter. Based on the marginal likelihood, they found that an ignorable model was more supported by the data than a nonignorable model. Then, they resorted to express uncertainty about ignorability by "centering" a nonignorable model on an ignorable one.

We use the NHIS data from the 1995 household survey to illustrate our method. In the NHIS the secondary sampling units are segments, and on average each segment includes about 4-12 households, and all the sample households in the segment are interviewed for core questions. We use the number of doctor visits by an entire household in the past year. We use the nine states with 8-12% nonrespondents (zip codes are shown in Table 1). We note that Colorado (CO), District of Columbia (DC) and Delaware (DE) are the three states different from the others: DC and DE have the least data and the highest proportions of nonrespondents, DE has the largest proportion of households with doctor visits, and CO has the smallest proportion of households with doctor visits.

The Bayesian method is discussed as a possible alternative to ratio estimation. In Section 2 we show that one cannot express uncertainty about ignorability by using a probabilistic mixture of an ignorable and a nonignorable model. In Section 3 we introduce a parameter which centers the nonignorable model on the ignorable model. An empirical study used to compare inference from the ignorable, nonignorable and expansion model is described in Section 4. Finally, Section 5 has concluding remarks.

2. IGNORABLE AND NONIGNORABLE MODELS

Let y_{ij} and r_{ij} be the characteristic and response variables for the j^{th} household in the i^{th} area, $i = 1, \ldots, \ell$, $j = 1, \ldots, n_i$. Also let A_{ij} denote the event that household j in area i has at least one doctor visit. Then if A_{ij} occurs, $y_{ij} = 1$; otherwise $y_{ij} = 0$. Let A^*_{ij} denote the event that household j in area i is a respondent. Then if A^*_{ij} occurs, $r_{ij} = 1$; otherwise $r_{ij} = 0$. The ignorable nonresponse model is given by

$$y_{ij} \mid p_i \overset{iid}{\sim} \text{Bernoulli } (p_i), \; r_{ij} \mid \pi_i \overset{iid}{\sim} \text{Bernoulli } (\pi_i),$$

$$p_i \mid \mu_{11}, \tau_{11} \overset{iid}{\sim} \text{Beta } (\mu_{11}\tau_{11}, (1-\mu_{11})\tau_{11}), \quad (1)$$

$$\pi_i \mid \mu_{12}, \tau_{12} \overset{iid}{\sim} \text{Beta } (\mu_{12}\tau_{12}, (1-\mu_{12})\tau_{12}). \quad (2)$$

The assumptions in (1) and (2) express similarity among areas. The nonignorable nonresponse model is given by

$$y_{ij} \mid p_i \overset{iid}{\sim} \text{Bernoulli } (p_i),$$

$$r_{ij} \mid y_{ij} = s-1, \; \pi_{is} \overset{iid}{\sim} \text{Bernoulli } (\pi_{is}), \quad s = 1, 2,$$

$$p_i \mid \mu_{21}, \tau_{21} \overset{iid}{\sim} \text{Beta } (\mu_{21}\tau_{21}, (1-\mu_{21})\tau_{21}),$$

$$\pi_{is} \mid \mu_{2,s+1}, \tau_{2,s+1} \overset{iid}{\sim}$$
$$\text{Beta } (\mu_{2,s+1}\tau_{2,s+1}, (1-\mu_{2,s+1})\tau_{2,s+1}), \quad s = 1, 2.$$

Again we express similarity among the areas. This similarity helps when the weakly identified parameters like the π_{i1} and π_{i2} are estimated. Note that we can obtain the ignorable model from the nonignorable one by taking $\pi_{i1} = \pi_{i2}$.

While Stasny (1991) assumed that the hyperparameters are fixed but unknown (estimated by a maximum likelihood procedure), we include uncertainty in the estimation of these hyperparameters by assuming

$$\mu_{rs} \sim \text{Beta } (1, \; 1) \quad \text{and } \tau_{rs} \sim \; \Gamma \; (\eta^{(0)}_{rs}, \nu^{(0)}_{rs})$$

where for the ignorable model $r = 1$, $s = 1, 2$ and for the nonignorable model $r = 2$, $s = 1, 2, 3$. In our case the $\eta^{(0)}_{rs}$ and $\nu^{(0)}_{rs}$ are to be specified (see Nandram and Choi, 2000).

The ignorable model is a complete Bayesian specification in which the probability for a household response does not depend on its characteristic, but the nonignorable model has the specification in which the probability for a household response does depend on its characteristic. This is the selection model, and there is a bivariate probability mass function on (y_{ij}, r_{ij}). It is convenient to order the labels so that the respondents are first and the nonrespondents are second.

The parameters of interest are (p_i, δ_i) where for the ignorable model $\delta_i = \pi_i$ and for the nonignorable model $\delta_i = \pi_{i2}p_i + \pi_{i1}(1-p_i)$. That is, $\delta_i = Pr(r_{ij} = 1)$, the probability household j responds in area i. We also let $r_i = \sum_{j=1}^{n_i} r_{ij}$ and $y_i = \sum_{j=1}^{r_i} y_{ij}$.

We note that $z_i = \sum_{j=r_i+1}^{n_i} y_{ij}$ is a latent variable, the unknown number of households with the characteristic for the nonrespondents. (We introduce the z_i into our procedure because they simplify the computations and, if interest is on the finite population proportion, we need to understand their distributions.) Then, the number of households without the characteristic is $n_i - r_i - z_i$ among the nonrespondents. Note that we can draw a tree diagram which shows the bivariate probability mass function of (y_{ij}, r_{ij}), and that π_{i1}, π_{i2} and p_i are on the path of the tree containing the observed data. Thus, the selection model automatically incorporates the prior and posterior properties of the Z_i.

The ignorable model is fitted by a direct application of the algorithm of Nandram (1998). For our Metropolis-Hastings algorithms on either the ignorable model or the nonignorable model, there was convergence after 1000 iterations, and taking every tenth iterate, provides 1000 iterates which they used for obtain estimate of (p_i, π_i, δ_i).

The marginal likelihood is a natural method to compare the ignorable and the nonignorable models because the ratio of the marginal likelihoods is the Bayes factor (e.g., Kass and Raftery, 1995) which measures the evidence provided by the data for one model relative to the other. Note that one of the objectives is to find the posterior probability that the ignorable model holds given some prior belief about its plausibility.

For either model we use importance sampling to compute the marginal likelihood. The logartithm of the marginal likelihood for the ignorable (noignorable) model is -62.821 (-74.203) with a numerical standard error of 0.007 (0.045) giving a Bayes factor of 11.383 with a numerical standard error of 0.046. Assuming equal prior probabilities on the two models, the posterior probability that the ignorable model holds is approximately 1. Thus, one should not express uncertainty about ignorability by a mixture of the ignorable model and nonignorable model (discrete model expansion). For the NHIS data the ignorable model is simply dominant. Thus, Bayesian model averaging (Hoeting, Madigan, Raftery, and Volinsky, 1999) is not appropriate. However, for the NHIS data one should not discard the nonignorable model because the ignorable model does not contain a component to represent an informative missing data mechanism (see Forster and Smith 1998 for a similar result).

Thus, for the NHIS data one must be careful not to select the ignorable model since it is believed that there are many other models. One needs to "search" for an appropriate model. A sensible strategy is to embed the ignorable model in a larger model through a continuous model expansion (Draper, 1995).

3. MODEL WITH UNCERTAINTY ABOUT IGNORABILITY

We expand the ignorable model to incorporate uncertainty about ignorability by using a centering parameter γ_i for area i. The key idea is to take $\pi_{i2} = \gamma_i \pi_{i1}$ in the nonignorable model. We will call this model the expansion model, and later the gamma model.

3.1 Expansion Model

The expansion model for nonignorable nonresponse is, $j = 1, \ldots, n_i$, $i = 1, \ldots, \ell$,

$$y_{ij} \mid p_i \overset{iid}{\sim} \text{Ber}(p_i), \quad r_{ij} \mid \pi_i, y_{ij} = 0 \overset{iid}{\sim} \text{Ber}(\pi_i),$$

$$r_{ij} \mid \pi_i, \gamma_i, y_{ij} = 1 \overset{iid}{\sim} \text{Ber}(\gamma_i \pi_i), \quad 0 < \gamma_i \pi_i < 1.$$

Here γ_i is the ratio of the odds of success among respondents to the odds of success among all households for the i^{th} area. Thus, if the odds ratio is one, there is no difference between respondents and nonrespondents and if the odds ratio is smaller (larger) than one, there is a smaller (larger) proportion of successes among the respondents than the respondents. The parameter γ_i describes the extent of nonignorability of the response mechanism for area i. It is through the γ_i that we incorporate uncertainty about ignorability.

Now the parameters of interest are $(p_i, \delta_i, \gamma_i)$ where $\delta_i = \pi_i \{\gamma_i p_i + (1 - p_i)\}$, is the probability that a household in the population responds in area i. These parameters are expected to vary across the areas, but we might believe they share an effect (i.e., they come from the same stochastic process). With a belief that all the areas are similar we take $(p_i, \delta_i, \gamma_i)$ to have a common distribution.

For the p_i,we take

$$p_i \mid \mu_1, \tau_1 \overset{iid}{\sim} \text{Beta}(\mu_1 \tau_1, (1 - \mu_1)\tau_1), \quad i = 1, \ldots, \ell$$

independent of (π_i, γ_i) which, in turn, are jointly independent over i with

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

$$\gamma_i \mid \nu \overset{iid}{\sim} \Gamma(\nu, \nu), \ 0 < \gamma_i < 1/\pi_i, \ 0 < \pi_i < 1.$$

Therefore, the joint prior density $p(\pi_i, \gamma_i \mid \mu_2, \tau_2, \nu)$ for (π_i, γ_i) is given by,

$$\nu \gamma_i^{\nu-1} exp(-\nu \gamma_i) \ \frac{\pi_i^{\mu_2 \tau_2 - 1} \ (1 - \pi_i)^{(1-\mu_2)\tau_2 - 1}}{B(\mu_2 \tau_2, \ (1 - \mu_2)\tau_2) \ \ I_i(\mu_2, \tau_2, \nu)} \tag{3}$$

where $B(u, v)$ is the beta function, $I_i(\mu_2, \tau_2, \nu)$ is

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

and $f_1(\pi_i, \phi_i \mid \nu, \mu_2, \tau_2)$ is

$$\nu \phi_i^{\nu-1} \frac{\pi_i^{\mu_2 \tau_2 - 1} \ (1 - \pi_i)^{(1-\mu_2)\tau_2 - 1}}{B(\mu_2 \tau_2, \ (1 - \mu_2)\tau_2}, $$

$0 < \pi_i, \ \phi_i < 1, \ \nu > 0$. The ignorable model is a special case of the expansion model with $\gamma_i = 1$. Also note that if the γ_i were not bounded above, $E(\gamma_i \mid \nu) = 1$ and $Var(\gamma_i \mid \nu) = 1/\nu$ (i.e., we have attempted to center the expansion model on the ignorable model).

The hyperparameters are in turn specified to be independent with proper prior densities

$$\nu \sim \Gamma(\eta_3^{(0)}, \nu_3^{(0)}), \qquad \mu_1, \mu_2 \overset{iid}{\sim} Beta(1, 1)$$

and

$$\tau_1 \sim \Gamma(\eta_1^{(0)}, \nu_1^{(0)}), \qquad \tau_2 \sim \Gamma(\eta_2^{(0)}, \nu_2^{(0)})$$

where $(\eta_s^{(0)}, \eta_s^{(0)})$, $s = 1, 2, 3$ are to be specified. Letting $\Omega = (\mu_1, \mu_2, \tau_1, \tau_2, \nu)$, then the joint prior for Ω is $p(\Omega) = p(\mu_1)p(\mu_2)p(\tau_1)p(\tau_2)p(\nu)$ where $p(\mu_1)$, $p(\mu_2)$, $p(\tau_1)$, $p(\tau_2)$ and $p(\nu)$ are the corresponding prior densities.

Then, letting $Z = \{z : z_i = 0, \ldots, n_i - r_i, \ i = 1, \ldots, \ell\}$, the likelihood function is proportional to $f(\underset{\sim}{y}, \underset{\sim}{r} \mid \underset{\sim}{p}, \underset{\sim}{\gamma}, \underset{\sim}{\pi})$ where $f(\underset{\sim}{y}, \underset{\sim}{r} \mid \underset{\sim}{p}, \underset{\sim}{\gamma}, \underset{\sim}{\pi}) = \sum_{\underset{\sim}{z}: \ \underset{\sim}{z} \in Z} f(\underset{\sim}{y}, \underset{\sim}{r}, \underset{\sim}{z} \mid \underset{\sim}{p}, \underset{\sim}{\gamma}, \underset{\sim}{\pi})$ and $f(\underset{\sim}{y}, \underset{\sim}{r}, \underset{\sim}{z} \mid \underset{\sim}{p}, \underset{\sim}{\gamma}, \underset{\sim}{\pi})$ is given by, $K_i = \begin{pmatrix} n_i \\ r_i \end{pmatrix} \begin{pmatrix} r_i \\ y_i \end{pmatrix} \begin{pmatrix} n_i - r_i \\ z_i \end{pmatrix}$,

$$\prod_{i=1}^{\ell} \left\{K_i (\gamma_i \pi_i p_i)^{y_i} (\pi_i (1 - p_i))^{r_i - y_i}\right.$$

$$\left.\times \quad ((1 - \gamma_i \pi_i) p_i)^{z_i} \quad ((1 - \pi_i)(1 - p_i))^{n_i - r_i - z_i}\right\}.$$

By Bayes' theorem the joint posterior density of the parameters follow easily. But it is convenient to make the transformation $\phi_i = \gamma_i \pi_i$, for $i = 1, \ldots, \ell$ with π_i, p_i and z_i untransformed. Now let $A_i = y_i + z_i + \mu_1 \tau_1$, $B_i = n_i - y_i - z_i + (1-\mu_1)\tau_1$, $C_i = r_i - y_i + \mu_2 \tau_2$ and $D_i = n_i - r_i - z_i + (1 - \mu_2)\tau_2$. Then, the joint posterior density of all the parameters $(\Omega, \underset{\sim}{z}, \underset{\sim}{p}, \underset{\sim}{\pi}, \underset{\sim}{\phi})$, given $\underset{\sim}{y}$ and $\underset{\sim}{r}$, $f(\Omega, \underset{\sim}{p}, \underset{\sim}{\pi}, \underset{\sim}{\phi}, \underset{\sim}{z} \mid \underset{\sim}{y}, \underset{\sim}{r})$, is we have proportional to

$$p(\Omega) \prod_{i=1}^{\ell} \left\{ \binom{n_i - r_i}{z_i} \frac{p_i^{A_i - 1} (1-p_i)^{B_i - 1}}{B(\mu_1 \tau_1, (1-\mu_1)\tau_1)} \right.$$

$$\times \; \nu \phi_i^{y_i + \nu - 1} (1-\phi_i)^{z_i}$$

$$\left. \times \; \frac{\pi_i^{C_i - 1} (1-\pi_i)^{D_i - 1}}{B(\mu_2 \tau_2, (1-\mu_2)\tau_2)} \frac{\{\pi_i^{-1} exp(-\phi_i/\pi_i)\}^{\nu}}{I_i(\mu_2, \tau_2, \nu)} \right\},$$

$0 < \phi_i, \; \pi_i < 1$. We use this posterior density to make inference about p_i, δ_i and γ_i.

3.2 Computation for Expansion Model

Because the posterior density is not accessible directly, we use MCMC methods to obtain samples which permit an inference.

For the NHIS data we take $\eta_1^{(0)} = 2.52$, $\nu_1^{(0)} = 0.05$, $\eta_2^{(0)} = 2.49$, $\nu_2^{(0)} = 0.18$, $\eta_3^{(0)} = \nu_3^{(0)} = .001$ and for the NCS data $\eta_1^{(0)} = 2.77$, $\nu_1^{(0)} = 0.05$, $\eta_2^{(0)} = 2.98$, $\nu_2^{(0)} = 0.05$, $\eta_3^{(0)} = \nu_3^{(0)} = .001$. We obtained these values by setting all of them equal .001, ran the algorithm, and then fit gamma densities to the random iterates for ν, τ_1 and τ_2. We also found that posterior inference about $(p_i, \gamma_i, \delta_i)$ was unchanged with these new values. This ensures posterior propriety and stability.

We marginalize out the parameters (p_i, π_i, ϕ_i), $i = 1, \ldots, \ell$ from the joint posterior density to obtain the posterior density $f(\Omega, \underset{\sim}{z} \mid \underset{\sim}{y}, \underset{\sim}{r})$. Then, we obtain samples from $f(\Omega, \underset{\sim}{z} \mid \underset{\sim}{y}, \underset{\sim}{r})$ using a Metropolis-Hastings algorithm and a sample importance resampling (SIR) algorithm (e.g., Smith and Gelfand 1992).

First, we define

$$f_a(\Omega, \underset{\sim}{z} \mid \underset{\sim}{y}, \underset{\sim}{r})$$

$$\propto \; p(\Omega) \prod_{i=1}^{\ell} \left\{ \binom{n_i - r_i}{z_i} \left\{ \frac{B(A_i, B_i)}{B(\mu_1 \tau_1, (1-\mu_1)\tau_1)} \right\} \right.$$

$$\left. \times \; \left\{ \frac{B(C_i, D_i)}{B(\mu_2 \tau_2, (1-\mu_2)\tau_2)} \right\} \nu \; B(y_i + \nu, z_i + 1) \right\}.$$

Then the marginal density $f(\Omega, \underset{\sim}{z} \mid \underset{\sim}{y}, \underset{\sim}{r})$ is proportional to

$$f_a(\Omega, \underset{\sim}{z} \mid \underset{\sim}{y}, \underset{\sim}{r}) \prod_{i=1}^{\ell} \{R_{z_i}(\mu_2, \tau_2, \nu)\}$$

$$R_{z_i}(\mu_2, \tau_2, \nu) = \frac{I_{z_i}(\mu_2, \tau_2, \nu)}{I_i(\mu_2, \tau_2, \nu)}, \tag{4}$$

where

$$I_{z_i}(\mu_2, \tau_2, \nu) = \int_0^1 \int_0^1 \{\pi_i^{-1} exp(-\phi_i/\pi_i)\}^{\nu} f_2(.) d\pi_i d\phi_i,$$

$I_i(\mu_2, \tau_2, \nu)$ is given in (3) and $f_2(.) = f_2(\pi_i, \phi_i \mid \nu, \mu_2, \tau_2, z_i, y_i, r_i)$ is

$$\frac{\phi_i^{y_i + \nu - 1}(1-\phi_i)^{z_i}}{B(y_i + \nu, \; z_i + 1)} \frac{\pi_i^{C_i - 1}(1-\pi_i)^{D_i - 1}}{B(C_i, \; D_i)}.$$

Observe that $R_{z_i}(\mu_2, \tau_2, \nu)$ is the ratio of the expectations of $\{\pi_i^{-1} exp(-\phi_i/\pi_i)\}^{\nu}$ over $f_2(\pi_i, \phi_i \mid \nu, \mu_2, \tau_2, z_i, y_i, r_i)$ and $f_1(\pi_i, \phi_i \mid \nu, \mu_2, \tau_2)$ in the numerator and denominator respectively.

Samples from $f_a(\Omega, \underset{\sim}{z} \mid \underset{\sim}{y}, \underset{\sim}{r})$ can be obtained by using the algorithm of Nandram (1998). We can obtain an observation ν from the conditional posterior density $p(\nu \mid \underset{\sim}{y}, \underset{\sim}{r}) \propto p(\nu) \prod_{i=1}^{\ell} \{ \; \nu \; B(y_i + \nu, \; z_i + 1)\}$. These samples are converted to samples from $f(\Omega, \underset{\sim}{z} \mid \underset{\sim}{y}, \underset{\sim}{r})$ using the SIR algorithm.

Once z_i, $i = 1, \ldots, \ell$ are obtained, we can draw p_i, π_i and ϕ_i from

$$p_i \mid y_i, r_i, z_i \overset{ind}{\sim} \text{Beta}\,(A_i, \; B_i)$$

and $g(\pi_i, \phi_i \mid \Omega, y_i, r_i, z_i)$ which is is proportional to

$$\{\pi_i^{-1} exp(-\phi_i/\pi_i)\}^{\nu} \, g_a(\pi_i, \phi_i \mid \Omega, y_i, r_i, z_i, \nu),$$

for $0 \leq \pi_i, \phi_i \leq 1$, where

$$g_a(\pi_i, \phi_i \mid \Omega, y_i, r_i, z_i, \nu) = \frac{\phi_i^{y_i + \nu - 1}(1-\phi_i)^{z_i + 1 - 1}}{B(y_i + \nu, \; z_i + 1)}$$

$$\times \frac{\pi_i^{C_i - 1} \, (1-\pi_i)^{D_i - 1}}{B(C_i, \; D_i)}.$$

It is straightforward to draw samples p_i from above relation, but it is more difficult to draw samples π_i and ϕ_i from equation $g(\pi_i, \phi_i \mid \Omega, y_i, r_i, z_i)$.

We finally obtain a sample $(\pi_i^{(h)}, \phi_i^{(h)}, \gamma_i^{(h)})$ by taking $\gamma_i^{(h)} = \pi_i^{(h)} \phi_i^{(h)}$, $h = 1, ..., M$. Inference can now be made in the standard way. We drew 11,000 iterates, used a "burn in" of 1000, picked every tenth thereafter to obtain 1000 iterates. We used the trace plots and the autocorrelations to confirm that the quality of the sample is good (Cowles and Carlin 1996).

We have computed the logarithm of the marginal likelihood for the expansion model. As compared with the values in Section 2 the logarithm of the marginal likelihood is -53.202 with a numerical standard error of 0.332. We are pleased that the expansion model dominates the ignorable and the nonignorable models individually.

4. AN EMPIRICAL STUDY

We apply our method to the data from the NHIS for the nine states to compare the three models and to study the effects of the centering parameters γ_i on ignorability and on inference about p_i and δ_i.

First, using the iterates from the Metropolis-Hastings sampler we drew the posterior densities of the γ_i represented by the histograms (not shown) which are all unimodal, and mostly skewed to the left. This is expected since the γ_i are bounded above by π_i^{-1} and the π_i are close to unity. For the NHIS data the left tails are very thin, and the posterior probabilities to the left of 1 are all small.

Table 1: Posterior mean (PM) and standard deviation (PSD), numerical standard error (NSE), 95% credible interval for γ and $\rho = Pr(\gamma \leq 1 \mid \tilde{y}, \tilde{r})$ for NHIS data of nine states with nonresponse rates 8 – 12%.

State	PM	PSD	NSE	Interval	ρ
CO	1.104	.066	.015	(0.906, 1.184)	.074
DC	1.115	.060	.014	(0.981, 1.221)	.036
DE	1.114	.073	.012	(0.935, 1.229)	.072
FL	1.111	.040	.008	(0.986, 1.152)	.031
LA	1.105	.040	.008	(1.008, 1.164)	.022
MD	1.119	.044	.009	(1.004, 1.178)	.022
NY	1.109	.043	.009	(0.985, 1.150)	.030
SC	1.105	.042	.009	(0.990, 1.172)	.027
WV	1.109	.047	.008	(0.995, 1.188)	.029

We consider the posterior densities of the γ_i even further by looking at the posterior mean (PM), standard deviation (PSD) and 95% credible interval. We also present the numerical standard error (NSE) for the Monte Carlo computation. (We use the batch means method with batch length of 25 for the 1000 "good" iterates.) We present these quantities for the NHIS data in Table 1. The NSE indicates reasonable Monte Carlo error, but the NSEs for Colorado, District of Columbia and Delaware are a bit larger than the others reflecting the small sample sizes for these states. Except for LA and MD all the credible intervals contain 1, and so one might consider the nonresponse mechanism for each state to be ignorable.

We were extremely surprised that the confidence intervals hardly provide any evidence of nonignorability. Thus, we calculated $\rho_i = Pr(\gamma_i \leq 1 \mid \tilde{y}, \tilde{r})$ which we present in the last column of Table 1. For the NHIS all the ρ_i are smaller than .10 and seven of them are smaller than .04. In addition, we looked at the box plots of the estimated posterior densities of γ_i from the iterates for both the NHIS. All the boxes are above 1.0, and most of them have medians above 1.1. Thus, in all nine states, there are substantial evidence for nonignorability contrary to the evidence provided by the credible intervals.

Next, in Table 2 we compare 95% credible intervals of the p_i and δ_i for the ignorable, nonignorable and the expansion models. Generally, there are differences among the three models for each data set.

First, consider the p_i for the NHIS data. The intervals based on the ignorable model are mostly contained by the intervals based on the nonignorable model with the lower bounds mostly similar. The intervals for the expansion model overlap on the left of the intervals for the ignorable model, making them considerably different. Now consider the δ_i. For the three models the intervals for FL, LA, NY, SC and WV are mostly very similar. There are differences for the others notably DC and DE.

In general, when the nonignorable model is used, the p_i might be too large. The expansion model (the best among the three models) is attractive because it fixes this problem. When the γ_i are larger than 1, the proportion of successes among the respondents is larger than that among the nonrespondents. The nonignorable model does not have the γ_i and so its absence makes the proportion of successes among the respondents smaller than that among the nonrespondents.

5. CONCLUDING REMARKS

We have studied nonignorable nonresponse for inference about (a) the proportion with a characteristic and (b) the proportion responding to the survey when there are data from similar areas. We have shown through a full Bayesian approach that it is possible to circumvent some of the issues associated with estimability especially for parameters with little relation to the data. This is accomplished by using a joint distribution on the success and response indicators as well as pooling data across similar areas.

We have argued that one should not incorporate uncertainty about ignorability by using a mixture of an ignorable model and a nonignorable model (i.e., discrete model expansion). Then, in our major contribution we have shown that it is sensible to include uncertainty about ignorability through a parameter which centers a nonignorable model on an ignorable one (i.e., continuous model expansion).

There are differences in inference about p_i and δ_i for the three models. We have shown that it is plau-

Table 2: 95% credible intervals for p_i and δ_i from the NHIS data by model

State	p_i	δ_i
a. Ignorable		
CO	(.222, .303)	(.871, .918)
CO	(.276, .330)	(.896, .917)
DC	(.285, .346)	(.895, .919)
DE	(.288, .351)	(.896, .919)
FL	(.297, .331)	(.901, .917)
LA	(.293, .342)	(.898, .919)
MD	(.296, .346)	(.897, .918)
NY	(.293, .324)	(.900, .916)
SC	(.290, .344)	(.898, .919)
WV	(.285, .344)	(.897, .919)
b. Nonignorable		
CO	(.266, .348)	(.874, .919)
DC	(.295, .424)	(.802, .931)
DE	(.292, .447)	(.846, .944)
FL	(.295, .363)	(.901, .925)
LA	(.294, .372)	(.890, .936)
MD	(.298, .399)	(.877, .923)
NY	(.287, .348)	(.899, .922)
SC	(.299, .390)	(.889, .936)
WV	(.274, .399)	(.856, .939)
c. Expansion		
CO	(.222, .303)	(.871, .918)
DC	(.253, .388)	(.841, .937)
DE	(.228, .363)	(.828, .929)
FL	(.273, .317)	(.900, .921)
LA	(.261, .337)	(.890, .932)
MD	(.274, .342)	(.885, .925)
NY	(.267, .311)	(.900, .920)
SC	(.268, .343)	(.886, .935)
WV	(.246, .348)	(.864, .934)

NOTE: The expansion model is a nonignorable model centered on the ignorable model; p_i is the proportion of visits and δ_i is the proportion of respondents in i^{th} state for the population.

sible that the nonresponse for all the nine states in the NHIS data is nonignorable. In addition, the nonignorable model makes the p_i too large, but the expansion model corrects this problem. The expansion model is preferred (supported by the Bayes factor) because the parameters γ_i form a useful method to study uncertainty about ignorability, and to adjust for nonresponse bias.

Reference

Cowles, M. and Carlin, B. (1996), "Markov chain Monte Carlo diagnostics: A comparative review," *Journal of the American Statistical Association*, 91, 883-904.

Draper, D. (1995), "Assessment and propagation of model uncertainty" (with discussion), *Journal of the Royal Statistical Society*, Series B, 57, 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*, Series B, 60, 57-70.

Hoeting, J. A., Madigan, D., Raftery, A. E. and Volinsky, C. T. (1999), "Bayesian model averaging: A tutorial," (with discussion), *Statistical Science*, 14, 382-417.

Kass, R. and Raftery, A. (1996), "Bayes factors," *Journal of the American Statistical Association*, 90, 773-795.

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

Nandram, B. (1998), "A Bayesian Analysis of the Three-Stage Hierarchical Multinomial Model," *Journal of Statistical Computation and Simulation*, 61, 97-126.

Nandram, B. and Choi, J. W. (2000), "Bayesian inference for a proportion under nonignorable nonresponse," *Technical Report*, National Center for Health Statistics, 6525 Belcrest Road, Hyattsville, MD 20782.

Smith, A. F. M. and Gelfand, A. E. (1992), "Bayesian Statistics Without Tears: A Sampling-resampling Perspective," *The American Statistician*, 46, 84-88.

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.