Journal of the Korean Statistical Society (2001), **30***: 3, pp 435–451*

A Hierarchical Bayesian Model for Survey Data with Nonresponse

Geunshik Han[1]

Abstract

We describe a hierarchical bayesian model to analyze multinomial nonignorable nonresponse data. Using a Dirichlet and beta prior to model the cell probabilities, We develop a complete hierarchical bayesian analysis for multinomial proportions without making any algebraic approximation. Inference is sampling based and Markov chain Monte Carlo methods are used to perform the computations. We apply our method to the data on body mass index (BMI) and show the model works reasonably well.

Keywords: Latent variable; Metropolis-Hastings sampler; Nonignorable nonresponse; Selection approach

1. INTRODUCTION

The nonresponse rates in many surveys have been increasing steadily (De Heer 1999 and Groves and Couper 1998), making the nonresponse problem more important. For many surveys the responses are multilevel. We describe a hierarchical Bayesian model to study multinomial nonignorable nonresponses.

Rubin(1987) and Little and Rubin(1987) describe two types of models according to the ignorability of response. 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 response must be distinct (see Rubin 1976). All other models are nonignorable. Our model essentially incorporates both types.

Crawford, Johnson and Laird (1993) used nonignorable nonresponse models to analyze data from the Harvard Medical Practice Survey. Stasny, Kadane, and Fritsch (1998) used a Bayesian hierarchical model for the probabilities of voting guilty or not on a particular trial given various death-penalty beliefs in which the

[1] Division of Computer and Information Science , Hanshin University, Osan 447-791, Korea

views of nonrespondents may differ from those of respondents. Park and Brown (1994) used a pseudo-Bayesian method (Baker and Laird 1988), and Park (1998) used a method in which prior observations are assigned to both observed and unobserved cells to estimate the missing cells of a multi-way categorical table under nonignorable nonresponse.

Stasny (1991) used a hierarchical Bayesian model to study victimization in the National Crime Survey (NCS), and used the selection approach developed primarily to study sample selection problems (e.g., Heckman 1976 and Olson 1980). A related approach was given by Albert and Gupta (1985) for a single area in which an approximation has made to obtain a Bayesian approach.

Stasny (1991) used a Bayes emperical Bayes method (Deely and Lindley 1981) in which the hyper-parameters are estimated using maximum likelihood methods and then assumed known. We extend this approach in two directions. First we consider multinomial data and second we provide a full Bayesian analysis.

It is possible to incorporate prior information about nonrespondents, and the Bayesian method is appropriate for the analysis of nonignorable nonresponse problems (Little and Rubin 1987 and Rubin 1987). The main difficulty is how to model the relationship between the respondents and nonrespondents.

The rest of the paper is organized as follows. In section 2 we discuss the Bayesian model for nonignorable nonresponses. In particular, a three-stage Bayesian hierarchical multinomial model is shown. In section 3 we describe the empirical study to assess the performance of our model. Finally, section 4 has conclusion remarks.

2. METHODOLOGY FOR HIERARCHICAL MULTINOMIAL MODEL

In this section we describe the Bayesian model for both ignorable and nonignorable nonresponse. For each subgroup(e.g., age, race, sex), an individual k in area i belongs to one of J categories, then for k^{th} individual with category j in area i, characteristic variable can be defined as follows,

$$\mathbf{x}_{ik} = (x_{i1k},\ ..., x_{ijk}, ...,\ x_{iJk})',\ i = 1, ..., c;\ j = 1, ..., J;\ k = 1, ..., n_i,$$

where each $x_{ijk} = 0$ or 1 and $\sum_{j=1}^{J} x_{ijk} = 1$.

The response variable, y_{ijk} is defined for each subgroup

$$y_{ijk} = \begin{cases} 1, & \text{if individual } k \text{ belonging to BMI level } j \text{ in county } i \text{ responded} \\ 0, & \text{if individual } k \text{ belonging to BMI level } j \text{ in county } i \text{ did not respond} \end{cases}$$

We use a probabilistic structure to model the $\mathbf{x}_{ik}$ and y_{ijk}.

2.1. Ignorable Versus Nonignorable Nonresponse

For ignorable nonresponse, response mechanism does not depend on characteristic since this is well supported by the data, hence we take

$$\mathbf{x}_{ik} \mid \mathbf{p}_i \overset{iid}{\sim} \text{Multinomial}\,(1, \mathbf{p}_i) \tag{1}$$

and

$$y_{ijk} \mid \pi_i \overset{iid}{\sim} \text{Bernoulli}\,(\pi_i). \tag{2}$$

At the second stage we take

$$\mathbf{p}_i \mid \boldsymbol{\mu}_1, \tau_1 \overset{iid}{\sim} \text{Dirichlet}\,(\boldsymbol{\mu}_1 \tau_1) \tag{3}$$

$$\pi_i \mid \mu_{21}, \tau_{21} \overset{iid}{\sim} \text{Beta}\,(\mu_{21}\tau_{21}, (1-\mu_{21})\tau_{21}), \tag{4}$$

where $p(\mathbf{p}_i \mid \boldsymbol{\mu}_1, \tau_1) = \frac{\prod_{j=1}^J p_{ij}^{\mu_{1j}\tau_1 - 1}}{D(\boldsymbol{\mu}_1 \tau_1)}$, $0 < p_{ij} < 1$, $\sum_{j=1}^J p_{ij} = 1$ and $\boldsymbol{\mu}_1 =$

$(\mu_{11}, \ldots, \mu_{1J})'$ with $D(\boldsymbol{\mu}_1 \tau_1) = \frac{\prod_{j=1}^J \Gamma(\mu_{1j}\tau_1)}{\Gamma(\tau_1)}$, $0 < \mu_{1j} < 1$, $\sum_{j=1}^J \mu_{1j} = 1$. Assumptions (3) and (4) express similarity among the areas.

For nonignorable nonresponse, we use the same model (1) for characteristic variable as that of ignorable model, but for the response variable we take

$$y_{ijk} \mid \mathbf{x}_{ik} = (x_{i1k}, ..., x_{iJk}), \pi_{ij} \overset{iid}{\sim} \text{Bernoulli}\,(\pi_{ij}), \tag{5}$$

where $x_{ijk} = 1$, $x_{ij'k} = 0$, $j \neq j'$ for j, $j' = 1, 2, ..., J$. At the second stage we also take

$$\mathbf{p}_i \mid \boldsymbol{\mu}_3, \tau_3 \overset{iid}{\sim} \text{Dirichlet}\,(\boldsymbol{\mu}_3 \tau_3), \tag{6}$$

where $\boldsymbol{\mu}_3 = (\mu_{31}, \mu_{32}, ..., \mu_{3J})'$.

$$\pi_{ij} \mid \mu_{4j}, \tau_{4j} \overset{iid}{\sim} \text{Beta}\,(\mu_{4j}\tau_{4j}, (1-\mu_{4j})\tau_{4j}),\ \text{j} = 1, ..., \text{J} \tag{7}$$

Like (3) and (4), the assumptions (6) and (7) express similarity among the counties. However the response parameters π_{ij} are weakly identifiable in this case. (7) helps in the estimation of the π_{ij}.

To ensure a full Bayesian analysis, at the third stage we take the prior for the hyper-parameters as follows. For the ignorable model
$\boldsymbol{\mu}_1 \sim$ Dirichlet $(1,1,...,1)$, $\mu_{21} \sim$ Beta $(1,1)$, $\tau_1 \sim \Gamma\ (\eta_1^{(0)}, \nu_1^{(0)})$, and $\tau_{21} \sim \Gamma\ (\eta_{21}^{(0)}, \nu_{21}^{(0)})$, where gamma density is given by $t \sim \Gamma\ (a,b)$ means $f(t) = b^a t^{a-1} e^{-bt}/\Gamma(a),\ t > 0$.
The corresponding part of the nonignorable model is
$\boldsymbol{\mu}_3 \sim$ Dirichlet $(1,1,...,1)$, $\mu_{4s} \sim$ Beta $(1,1)$,$\tau_3 \sim \Gamma\ (\eta_3^{(0)}, \nu_3^{(0)})$, and $\tau_{4s} \sim \Gamma\ (\eta_{4s}^{(0)}, \nu_{4s}^{(0)}), j = 1,...,J$.

The hyper-parameters $\eta_3^{(0)}$, $\nu_3^{(0)}$, $\eta_{4s}^{(0)}$ and $\nu_{4s}^{(0)}$ $s = 1,...,J$ are to be specified. Let r_i be the number of respondents in county i and y_{ij} the number of respondents for j th BMI level in county i. Then r_i and y_{ij} are random variables, $n_i - r_i$ is the number of nonrespondents, since the number of respondents for jth BMI level for the nonrespondents is unknown, we denote them by the latent variables z_{ij} (see the tree diagram in Figure 1)

The likelihood function for ignorable nonresponse is

$$f(\mathbf{y}, \mathbf{r} \mid \mathbf{p}_i, \pi_i) = \prod_{i=1}^{c} \left\{ \binom{n_i}{r_i} \pi_i^{r_i} (1-\pi_i)^{n_i - r_i} \right\} \prod_{i=1}^{c} \left\{ \binom{r_i}{y_{i1}, .., y_{iJ}} \prod_{j=1}^{J} \left\{ p_{ij}^{y_{ij} + n_i - r_i} \right\} \right\}.$$

Here the likelihood functions for p_{ij} and π_i can be seperated out and using Bayes theorem the joint posterior density of all the parameters is expressed as

$$\begin{aligned} & f(\mathbf{p}, \boldsymbol{\pi}, \boldsymbol{\mu}_1, \tau_1, \mu_{21}, \tau_{21} \mid \mathbf{y}, \mathbf{r}) \\ & \propto \prod_{i=1}^{c} \left\{ \prod_{j=1}^{J} p_{ij}^{y_{ij} + n_i - r_i} \right\} \{ \pi_i^{r_i} (1-\pi_i)^{n_i - r_i} \} \\ & \times \frac{\prod_{j=1}^{J} p_{ij}^{\mu_{1j}\tau_1 - 1}}{D(\boldsymbol{\mu}_1 \tau_1)} \times \frac{\pi_i^{\mu_{21}\tau_{21} - 1} (1-\pi_i)^{(1-\mu_{21})\tau_{21} - 1}}{B(\mu_{21}\tau_{21}, (1-\mu_{21})\tau_{21})} \\ & \times \left\{ \tau_1^{\eta_1^{(0)} - 1} exp(-\nu_1^{(0)} \tau_1) \right\} \left\{ \tau_{21}^{\eta_{21}^{(0)} - 1} exp(-\nu_{21}^{(0)} \tau_{21}) \right\}. \qquad (8) \end{aligned}$$

Similarly the likelihood function for nonignorable nonresponse model is

$$\begin{aligned} f(\mathbf{y},\mathbf{r} \mid \mathbf{p}_i,\boldsymbol{\pi}_i,\mathbf{z}) &= \prod_{i=1}^{c}\left\{\binom{n_i}{r_i}\binom{r_i}{y_{i1},..,y_{iJ}}\binom{n_i-r_i}{z_{i1},..,z_{iJ}}\right. \\ &\times \left.\prod_{j=1}^{J}\{(\pi_{ij}p_{ij})^{y_{ij}}((1-\pi_{ij})p_{ij})^{z_{ij}}\}\right\}. \end{aligned}$$

And using Bayes' theorem the joint posterior density of all the parameters is

$$\begin{aligned} & f(\mathbf{p},\boldsymbol{\pi},\mathbf{z},\boldsymbol{\mu}_3,\tau_3,\boldsymbol{\mu}_4,\boldsymbol{\tau}_4 \mid \mathbf{y},\mathbf{r}) \\ & \propto \prod_{i=1}^{c}\left\{\binom{n_i-r_i}{z_{i1},\cdot\cdot,z_{iJ}}\prod_{j=1}^{J}(\pi_{ij}p_{ij})^{y_{ij}}((1-\pi_{ij})p_{ij})^{z_{ij}}\right. \\ & \times \left.\frac{\prod_{j=1}^{J}p_{ij}^{\mu_{3j}\tau_3-1}}{D(\boldsymbol{\mu}_3\tau_3)}\prod_{j=1}^{J}\left\{\frac{\pi_{ij}^{\mu_{4j}\tau_{4j}-1}(1-\pi_{ij})^{(1-\mu_{4j})\tau_{4j}-1}}{B(\mu_{4j}\tau_{4j},(1-\mu_{4j})\tau_{4j})}\right\}\right\} \\ & \times \left\{\tau_3^{\eta_2^{(0)}-1}exp(-\nu_2^{(0)}\tau_3)\right\}\prod_{j=1}^{J}\left\{\tau_{4j}^{\eta_{4j}^{(0)}-1}exp(-\nu_{4j}^{(0)}\tau_{4j})\right\}. \qquad (9) \end{aligned}$$

We consider inferences about $\mathbf{p}_i$ and probability of responding, $\delta_i = \sum_{j=1}^{J}\pi_{ij}p_{ij}$. Our plan is to obtain Metropolis-Hastings(MH) samplers to get sample from (8) and (9) and then to use these sample to make a posterior about $\mathbf{p}_i$ and δ_i.

2.2. Computations

We use Markov Chain Monte Carlo (MCMC) algorithm to obtain the posterior distribution of the $\mathbf{p}_i$ and $\boldsymbol{\pi}_i$. For the ignorable nonresponse model it is convenient to represent the posterior density function as

$$\begin{aligned} & f(\mathbf{p},\pi,\boldsymbol{\mu}_1,\tau_1,\mu_{21},\tau_{21} \mid \mathbf{y},\mathbf{r}) = \\ & \prod_{i=1}^{c}\{f_1(\mathbf{p}_i \mid \mathbf{y},\mathbf{r},\boldsymbol{\mu}_1,\tau_1)f_2(\pi_i \mid \mathbf{y},\mathbf{r},\mu_{21},\tau_{21})\}\, f_3(\boldsymbol{\mu}_1,\tau_1,\mu_{21},\tau_{21} \mid \mathbf{y},\mathbf{r}) \end{aligned}$$

where $f_1(\cdot)$ is Dirichlet density with

$$\mathbf{p}_i \mid \mathbf{y}_i,r_i,\boldsymbol{\mu}_1,\tau_1 \overset{ind}{\sim} \mathrm{D}(\mathbf{y}_i+n_i-r_i+\boldsymbol{\mu}_1\tau_1),$$

$f_2(\cdot)$ is beta density which is

$$\pi_i \mid y_i,r_i,\mu_{21},\tau_{21} \overset{ind}{\sim} \mathrm{Beta}(r_i+\mu_{21}\tau_{21},\ n_i-r_i+(1-\mu_{21})\tau_{21})$$

and

$$\begin{aligned} & f_3(\boldsymbol{\mu}_1, \tau_1, \mu_{21}, \tau_{21} \mid \mathbf{y}, \mathbf{r}) \\ \propto & \prod_{i=1}^{c} \left\{ \frac{D(\mathbf{y}_i + n_i - r_i + \boldsymbol{\mu}_1 \tau_1)}{D(\boldsymbol{\mu}_1 \tau_1)} \right\} p(\boldsymbol{\mu}_1, \tau_1) \\ \times & \prod_{i=1}^{c} \left\{ \frac{B(r_i + \mu_{21}\tau_{21},\ n_i - r_i + (1-\mu_{21})\tau_{21})}{B(\mu_{21}\tau_{21}, (1-\mu_{21})\tau_{21})} \right\} p(\mu_{21}, \tau_{21}) \end{aligned}$$

where $p(\boldsymbol{\mu}_1, \tau_1)$ and $p(\mu_{21}, \tau_{21})$ are the joint prior distribution. f_1 and f_2 are obtained through the Gibbs kernel, while for f_3 we use the MH algorithm (see appendix 1) of Nandram (1998).

For nonignorable nonresponse model it is convenient to represent the posterior density function as

$$\begin{aligned} & f(\mathbf{p}, \boldsymbol{\pi}, \mathbf{z}, \boldsymbol{\mu}_3, \tau_3, \boldsymbol{\mu}_4, \boldsymbol{\tau}_4 \mid \mathbf{y}, \mathbf{r}) \\ = & \prod_{i=1}^{c} \left\{ \left\{ \prod_{j=1}^{J} f_j(\pi_{ij} \mid \mathbf{y}, \mathbf{r}, \mathbf{z}, \mu_{4j}, \tau_{4j}) \right\} f_{J+1}(\mathbf{p}_i \mid \mathbf{y}, \mathbf{r}, \mathbf{z}, \boldsymbol{\mu}_3, \tau_3) \right\} \\ \times & \ f_{J+2}(\boldsymbol{\mu}_3, \tau_3, \boldsymbol{\mu}_4, \boldsymbol{\tau}_4, \mathbf{z} \mid \mathbf{y}, \mathbf{r}), \end{aligned}$$

where $f_1(\cdot)$, ..., $f_J(\cdot)$ are beta densities with

$$\pi_{ij} \mid y_{ij}, r_{ij}, z_{ij}, \mu_{4j}, \tau_{4j} \overset{ind}{\sim} \text{Beta}(y_{ij} + \mu_{4j}\tau_{4j}, z_{ij} + (1-\mu_{4j})\tau_{4j}),$$

and $f_{J+1}(\cdot)$ is Dirichlet density with

$$\mathbf{p}_i \mid \mathbf{y}_i, \mathbf{z}_i, \boldsymbol{\mu}_3, \tau_3 \overset{ind}{\sim} \mathrm{D}(\mathbf{y}_i + \mathbf{z}_i + \boldsymbol{\mu}_3\tau_3),$$

and $f_{J+2}(\cdot)$ is

$$\begin{aligned} & f_{J+2}(\boldsymbol{\mu}_3, \tau_3, \boldsymbol{\mu}_4, \boldsymbol{\tau}_4, \mathbf{z} \mid \mathbf{y}, \mathbf{r}) \\ \propto & \prod_{i=1}^{c} \left\{ \begin{pmatrix} n_i - r_i \\ z_{i1}, .., z_{iJ} \end{pmatrix} \frac{D(\mathbf{y}_i + \mathbf{z}_i + \boldsymbol{\mu}_3\tau_3)}{D(\boldsymbol{\mu}_3\tau_3)} p(\boldsymbol{\mu}_3, \tau_3) \right. \\ \times & \left. \prod_{j=1}^{J} \frac{B(y_{ij} + \mu_{4j}\tau_{4j}, z_{ij} + (1-\mu_{4j})\tau_{4j})}{B(\mu_{4j}\tau_{4j}, (1-\mu_{4j})\tau_{4j})} p(\boldsymbol{\mu}_4, \boldsymbol{\tau}_4) \right\} \end{aligned}$$

where $p(\boldsymbol{\mu}_3, \boldsymbol{\tau}_3)$ and $p(\boldsymbol{\mu}_4, \boldsymbol{\tau}_4)$ are the joint prior distribution. Thus, $f_1, ..., f_{J+1}$ are obtained through the Gibbs kernel, while f_{J+2} is obtained using the MH algorithm (see appendix 1) of Nandram (1998).

We monitor the convergence of the MCMC using the Gelman and Rubin (1992) method that uses the analysis of variance technique to determine whether further iterations are needed. We found 500 iterations to be enough for the priors being considered. All the numerical results are obtained with 2,000 iterations after buen-in iterations and we took every second samples.

It is enough to wash out autocorrelation among the iterates and to have good jumping probabilities (0.25-0.50). For computation, first we set $\eta_1^{(0)}$, $\nu_1^{(0)}$, $\eta_{21}^{(0)}$, $\nu_{21}^{(0)}$, $\eta_3^{(0)}$, $\nu_3^{(0)}$, $\eta_{4s}^{(0)}$, $\nu_{4s}^{(0)}$, $s = 1, ..., J$, equal to 0. Then we ran our MH algorithm to obtain posterior samples of τ_1, τ_{21}, τ_3 and τ_{4s}, s=1,...,J. To ensure propriety of the posterior we estimate $\eta_1^{(0)}$, $\nu_1^{(0)}$, $\eta_{21}^{(0)}$, $\nu_{21}^{(0)}$, $\eta_3^{(0)}$, $\nu_3^{(0)}$, $\eta_{4s}^{(0)}$, $\nu_{4s}^{(0)}$ j=1,...,J, by fitting the gamma prior on the posterior samples for τ_1, τ_{21}, τ_3 and tau_{4s}, $s = 1, ..., J$. These values are shown in Table 1. Finally, with these proper priors we ran our algorithm to obtain posterior samples.

3. AN EMPIRICAL ANALYSIS

In this section we describe an empirical analysis using the National Health and Nutrition Examination Survey (NHANES) data to illustrate our methodology. The data for our illustration come from this survey, and was collected October 1988 and September 1994.

The NHANES consists of two parts, first part is the interview of the sampled person for their personal information and second part is the examination of those sampled. The persons from the sample of households were grouped into a number of subgroups depending on the age, race and sex. Some subgroups were sampled at different rates. Sampled persons were asked to come to station for physical examination. Those who did not come were visited by the examiner for the same purpose. Details of the NHANES sample design are available (Vital and Health Statistics, Series 2, Number 113 1992).

One of the variables of interest in the NHANES is body mass index (BMI), a convenient index of weight adjusted for height (Kg/m^2) that can be used to broadly categorize bodyweight within age-race-sex groups (Kuczmarski et al. 1997) as low body fat (level 1: BMI $<$ 20), healthy body fat (level 2: 20 $\leq$ BMI $<$ 25), hefty or unhealthy (level 3: BMI $\geq$ 25). We use this classification for the each of 8 age-race-sex groups.

The main reasons for the NHANES nonresponse are "not interested", "no time/work conflict", "concerns/suspicious", "don't bother me" and "health reasons". The nonresponse rate of the young age group is high, especially the parents, older mothers of a only child, were extremely protective of their babies, and would not allow them to leave home for physical examination. Such nonresponse might be nonrandom and hence require some special modeling.

Table 2 shows the number of respondents for each BMI level by age-race-sex group for 34 counties (population at least 500,000). The pattern of respondents differs greatly by age groups (young: age $<$ 45 years ; old: age $\geq$ 45 years). The nonresponse rate for old age group is negligible but that for young age group is high. Therefore We use a two part hierarchical model for the BMI data. The first part of the model is for the old individuals, where we apply a ignorable model. The second part of the model is for the young individuals, and is a nonignorable model.

We develop a methodology to analyze the three category BMI data by age, race and sex, although our methodology applies generally to any number of cells in several areas (counties in our application). Since p_{ij} are similar for each county, we take the weighted posterior mean of p_{ij},

$$q_j = \sum_{i=1}^{c} n_i p_{ij} / \sum_{i=1}^{c} n_i, j = 1, 2, 3$$

by age, race and sex for both young and old age group.

3.1. Empirical Analysis

First we perform a sensitivity analysis to access the specifications of $\eta^{(0)}$ and $\nu^{(0)}$. We compared four choices of hyper-priors $\Omega = (\eta^{(0)}, \nu^{(0)})$ to check its sensitivity to inference. For choice 1, we use $4 \times \eta^{(0)}$ and $4 \times \nu^{(0)}$. For choice 2, we use $\eta^{(0)}$ and $\nu^{(0)}$. For choice 3, we use $\hat{\eta}/4$ and $\hat{\nu}/4$ and fourth choice was 0 for both hyper-priors.

Table 3 shows the simulation results of sensitivity to inference of q_j for young age group. the point estimates and standard deviations of the proportion are very similar over the four choices of hyper-priors. Similarly, Table 4 shows the simulation results of q_j for old age group. The point estimates of male group are very similar over the four choices of hyper-priors, but there are some changes in the female group. The changes in point estimate for the female group are from 4Ω to Ω, but there are no substantial changes from Ω to 0Ω.

Standard deviations are increase when multiplier of Ω decrease for the female group, but there are no substantial changes for the male group. Generally nonignorable performs better than ignorable model, in other word nonignorable model is not sensitive to choices of hyper priors while ignorable model is possibly sensitive to choices of hyper priors.

95% credible intervals for the weighted posterior mean are shown in Table 5. For young age group, the weighted posterior mean is highest for q_1 (BMI level 1), and q_2 (BMI level 2) is lowest. the lower bounds for q_1 and q_3 are similar for young age group except white-male, and those for q_2 are similar except others-male group. For old age group, the weighted posterior mean is highest for q_3 (BMI level 3), and q_1 (BMI level 2) is lowest. Specifically q_1, q_2 are high and q_3 is low for white-female group.

Table 6 shows point estimates of the probability of responding $\delta_i = \sum_{j=1}^{J} \pi_{ij} p_{ij}$, and their 95% credible intervals with choices of Ω. The probabilities of responding for male are lower than that for female, and this trend remaining the same for three choices of Ω. If a similar survey is conducted in the future we should increase sample size by $1.25 = (1/.8)$time for male and $1.35 = (1/.74)$ time for female(e.g., if complete data are required from 1,000 households, the interviewer needs to contact 1,250 households for males).

4. CONCLUSION

We have discussed the problem of nonignorable nonresponse for the estimation of the proportions. We have extended the method of Stasny (1991) in two directions. First we consider multinomial data other than binomial data and second we study a full Bayesian analysis. We applied our methodology to the NHANES data. The MCMC method allowed us to assess the complex structure of the multinomial nonresponse estimation. Our empirical analysis indicate good performance for our model for this data.

For the NHANES there are substantial differences in the proportion of individuals in the 3 BMI levels for males versus females and young versus old. While we have shown that inference about BMI is not sensitive to prior specification, we might want use other prior densities for the Dirichlet or beta parameters(i.e., a uniform shrinkage prior).

It is feasible to use a nonignorable model that incorporates the extent of nonignorability.

APPENDIX 1

Metropolis-Hastings Samplers

For ignorable mode $(\boldsymbol{\mu}_1, \tau_1)$ and (μ_{21}, τ_{21}) are independent aposteriori with

$$p(\boldsymbol{\mu}_1, \tau_1 \mid \mathbf{y}, \mathbf{r}) \propto p(\boldsymbol{\mu}_1, \tau_1) \prod_{i=1}^{c} \left\{ \frac{D(\mathbf{y}_i + n_i - \mathbf{r}_i + \boldsymbol{\mu}_1 \tau_1)}{D(\boldsymbol{\mu}_1 \tau_1)} \right\} \tag{A.1}$$

and

$$p(\mu_{21}, \tau_{21} \mid \mathbf{y}, \mathbf{r}) \propto p(\mu_{21}, \tau_{21}) \prod_{i=1}^{c} \left\{ \frac{B(r_i + \mu_{21}\tau_{21}, r_i - y_i + (1 - \mu_{21})\tau_{21})}{B(\mu_{21}\tau_{21}, (1 - \mu_{21})\tau_{21})} \right\} \tag{A.2}$$

where $p(\boldsymbol{\mu}_1 \tau_1)$ and $p(\mu_{21}, \tau_{21})$ are the prior distributions. Samples can be obtained from each of (A.1) and (A.2) by using the algorithm of Nandram (1998).

For nonignorable model it is convenient to condition on $\mathbf{z}$ to obtain

$$p(\boldsymbol{\mu}_3, \tau_3 \mid \mathbf{z}, \mathbf{y}, \mathbf{r}) \propto p(\boldsymbol{\mu}_3, \tau_3) \prod_{i=1}^{c} \left\{ \frac{D(\mathbf{y}_i + \mathbf{z}_i + \boldsymbol{\mu}_3 \tau_3)}{D(\boldsymbol{\mu}_3 \tau_3)} \right\} \tag{A.3}$$

$$p(\mu_{4s}, \tau_{4s} \mid \mathbf{z}, \mathbf{y}, \mathbf{r}) \propto p(\mu_{4s}, \tau_{4s}) \prod_{i=1}^{c} \left\{ \frac{B(y_{ij} + \mu_{4s}\tau_{4s}, z_{ij} + (1 - \mu_{4s})\tau_{4s})}{B(\mu_{4s}\tau_{4s}, (1 - \mu_{4s})\tau_{4s})} \right\}, \tag{A.4}$$

where $p(\boldsymbol{\mu}_3, \tau_3)$, $p(\mu_{4s}, \tau_{4s})$,$(s = 1, ..., J)$ are the prior distributions and for given $\mathbf{z}$ (A.3) and (A.4) are independent.

$$\begin{aligned} p(z_{i1} = t_1, ..., z_{iJ} = t_J \mid \mathbf{y}, \mathbf{r}, \mu_{1s}, \tau_1, \mu_{3s}, \tau_{3s}, s = 1, ..., J) \\ = \omega_{it_1t_2...t_J} / \sum_{t_1=0}^{n_i - r_i} \cdots \sum_{t_J=0}^{n_i - r_i} \omega_{it_1t_2...t_J}, \end{aligned} \tag{A.5}$$

for $t = 0, 1, \ldots, n_i - r_i$.

$$\omega_{it_1t_2...t_J} = \begin{pmatrix} n_i - r_i \\ t_1, ..., t_J \end{pmatrix} D(\mathbf{y}_i + \mathbf{t}_i + \boldsymbol{\mu}_3 \tau_3) \prod_{j=1}^{J} B(y_{ij} + \mu_{4j}\tau_{4j},\ t_{ij} + (1 - \mu_{4j})\tau_{4j}).$$

We run the Metropolis-Hastings sampler by drawing a random deviate from each of (A.3), (A.4), and (A.5). It is easy to draw a random deviate from (A.5). Samples can be obtained from each of (A.3), (A.4) and (A.5) by using the algorithm of Nandram (1998).

Table 1: Estimates of η and ν for τ_3,τ_{41},τ_{42},τ_{43} for young age group and η and ν for τ_1 and τ_{21} for old age group

Age	Race	Sex	τ	η	ν	mean	std
Young	White	Male	τ_3	3.69795	.03554	104.050	54.108
			τ_{41}	2.34070	.07118	32.883	21.493
			τ_{42}	3.08505	.20091	15.355	8.742
			τ_{43}	2.68450	.16349	16.420	10.022
		Female	τ_3	4.19975	.03045	137.909	67.295
			τ_{41}	3.29355	.05880	56.014	30.865
			τ_{42}	2.48146	.07194	34.495	21.898
			τ_{43}	1.81866	.01664	109.308	81.054
	Others	Male	τ_3	4.94848	.06828	72.473	32.579
			τ_{41}	2.92190	.09637	30.321	17.738
			τ_{42}	3.15611	.16859	18.721	10.538
			τ_{43}	2.40424	.14709	16.345	10.542
		Female	τ_3	3.74506	.05515	67.908	35.091
			τ_{41}	3.08397	.03566	86.492	49.252
			τ_{42}	1.89257	.04859	38.951	28.313
			τ_{43}	2.34964	.11644	20.179	13.164
Old	White	Male	τ_1	4.40801	.00869	507.216	241.586
			τ_{21}	3.94073	.05246	75.123	37.840
		Female	τ_1	4.78810	.00834	574.409	262.506
			τ_{21}	4.38366	.01912	229.319	109.527
	Others	Male	τ_1	5.97146	.10666	55.986	22.911
			τ_{21}	4.37649	.03572	122.511	58.562
		Female	τ_1	3.29226	.00867	379.862	209.353
			τ_{21}	4.48822	.03641	123.285	58.194

Table 2: Number of individuals in each BMI level and number of nonrespondents by age, race and sex over all 34 counties

			BMI			
Age	Race	Sex	1	2	3	Nonresponse
Young	White	Male	1098	651	597	558
		Female	845	434	380	233
	Others	Male	1198	713	665	574
		Female	745	463	524	214
Old	White	Male	46	439	1014	3
		Female	51	223	365	4
	Others	Male	79	470	942	8
		Female	48	169	552	6

Note : BMI-level 1: < 20; level 2: $20 - 25$; level 3: > 25; Young < 45 : Old ≥ 45

Table 3: Sensitivity of q_j for choice of $\eta_3^{(0)}$,$\nu_3^{(0)}$,$\eta_{4s}^{(0)}$, and $\nu_{4s}^{(0)}$, $s = 1,..,4$ for young age group

Race	Sex	q_1	$std(q_1)$	q_2	$std(q_2)$	q_3	$std(q_3)$
(a) 4Ω							
White	Male	.428	.022	.216	.019	.356	.022
	Female	.476	.025	.232	.020	.292	.024
Others	Male	.419	.020	.212	.016	.369	.020
	Female	.434	.026	.185	.023	.381	.027
(b) Ω							
White	Male	.427	.022	.211	.020	.362	.025
	Female	.476	.026	.223	.024	.301	.031
Others	Male	.419	.020	.208	.017	.373	.022
	Female	.435	.025	.178	.026	.387	.029
(c) $\Omega/4$							
White	Male	.427	.022	.210	.021	.364	.027
	Female	.475	.026	.220	.026	.304	.034
Others	Male	.419	.020	.206	.018	.375	.024
	Female	.435	.025	.177	.028	.388	.029
(d) 0Ω							
White	Male	.426	.022	.209	.022	.365	.027
	Female	.474	.026	.218	.027	.308	.034
Others	Male	.419	.020	.205	.018	.376	.024
	Female	.435	.025	.177	.029	.388	.029

Note1 : $\Omega = \left(\eta_3^{(0)}, \nu_3^{(0)}, \eta_{41}^{(0)}, \nu_{41}^{(0)}, \eta_{42}^{(0)}, \nu_{42}^{(0)}, \eta_{43}^{(0)}, \nu_{43}^{(0)}\right)$.

Note2 : Nonignorable model applied for young age group

Table 4: Sensitivity of q_j for choice of $\eta_1^{(0)}$,$\nu_1^{(0)}$, $\eta_{21}^{(0)}$,$\nu_{21}^{(0)}$ for old age group

Race	Sex	q_1	$std(q_1)$	q_2	$std(q_2)$	q_3	$std(q_3)$
(a) 4Ω							
White	Male	.030	.005	.306	.018	.664	.018
	Female	.081	.002	.436	.004	.483	.004
Others	Male	.053	.011	.317	.017	.630	.018
	Female	.075	.005	.201	.004	.724	.006
(b) Ω							
White	Male	.031	.005	.292	.016	.677	.016
	Female	.063	.002	.443	.006	.494	.005
Others	Male	.053	.011	.316	.019	.631	.020
	female	.066	.012	.237	.018	.697	.019
(c) Ω/4							
White	Male	.031	.005	.293	.018	.676	.019
	Female	.073	.015	.359	.011	.568	.019
Others	Male	.053	.010	.317	.018	.630	.019
	Female	.065	.013	.221	.022	.714	.025
(d) 0Ω							
White	Male	.031	.005	.293	.020	.677	.019
	Female	.080	.014	.359	.031	.561	.033
Others	Male	.053	.010	.316	.019	.631	.019
	Female	.066	.015	.218	.026	.717	.029

Note1 : Ignorable model applied for old age group

Table 5: 95% credible intervals for the weighted posterior means, $q_j = \sum_{i=1}^{c} n_i p_{ij} / \sum_{i=1}^{c} n_i$ by age, race and sex for each age group

			95% credible interval		
Age	Race	Sex	q_1	q_2	q_3
Young	White	Male	(.382 .470)	(.174 .252)	(.314 .412)
		Female	(.425 .525)	(.171 .269)	(.243 .371)
	Others	Male	(.381 .455)	(.176 .241)	(.333 .419)
		Female	(.385 .482)	(.130 .230)	(.329 .442)
Old	White	Male	(.022 .041)	(.255 .326)	(.643 .710)
		Female	(.059 .068)	(.431 .451)	(.486 .505)
	Others	Male	(.035 .076)	(.282 .352)	(.592 .670)
		Female	(.040 .093)	(.206 .265)	(.661 .731)

Note1 : Nonignorable model applied for young respondents.
Note2 : Ignorable model applied for old respondents.

Table 6: 95 % credible intervals for probability of response for three choices of Ω

		4Ω		Ω		$\Omega/4$	
Race	Sex	δ_i $std(\delta_i)$	interval	δ_i $std(\delta_i)$	interval	δ_i $std(\delta_i)$	interval
W	M	.775(.016)	(.744 .805)	.769(.017)	(.735 .801)	.767(.018)	(.732 .799)
	F	.855(.017)	(.821 .886)	.855(.020)	(.810 .887)	.853(.022)	(.806 .887)
O	M	.786(.016)	(.752 .817)	.780(.018)	(.740 .813)	.778(.018)	(.739 .811)
	F	.880(.013)	(.854 .902)	.878(.015)	(.845 .903)	.876(.015)	(.838 .903)

Note : Race: W, White; O, Others; Sex: M, male; F, Female

ACKNOWLEDGEMENT

We thank Dr. Balgobin Nandram for his support on the MetroPolis-Hastings algorithm.

References

Albert, J. H. and Gupta, A. K. (1985). Bayesian Methods for Binomial Data with Applications to a Nonresponse Problem. *Journal of the American statistical Association*, **80**, 167-174.

Baker, S. G. and Laird, N. M. (1988). Regression analysis for categorical variables with Outcome Subject to Nonignorable nonresponse. *Journal of the American statistical Association*, **83**, 62-69.

Crawford, S. L., Johnson, W. G. and Laird, N. M. (1993). Bayes Analysis of Model-Based Methods for Nonignorable Nonresponse in the Harvard Medical Practice Survey (with discussions). *In Case Studies in Bayesian Statistics.* Gatsonis C., Hodges, J.S., Kass, R.E. and Sinpurwalla, N.D. (eds.) Springer-Verlag: New York, pp. 78-117.

De Heer, W. (1999). International Response Trends: Results of an International Survey. *Journal of Official Statistics*, **15**, 129-142.

Deely, J. J. and Lindley, D. V. (1981). Bayes Empirical Bayes. *Journal of the American Statistical Association*, **76**, 833-841.

Gelman, A. E. and Rubin D. (1992). Inference from Iterative Simulation Using Multiple Sequences. *Statistical Science*, **7**, 457-472.

Groves, R. M. and Couper, M. P. (1998). *Nonresponse in Household Interview Surveys.* New York: Wiley.

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.

Kuczmarski, R. J., Carrol, M. D., Flegal, K. M. and Troiano, R. P. (1997). Varying Body Mass Index cutoff points to describe overweight prevalence among U.S. adults: NHANES III (1988 to 1994). *Obesity Research*, **5**, 542-548.

Little R. J. A. and Rubin D. B. (1987). *Statistical Analysis with Missing Data.* New York: Wiley.

Olson, R. L. (1980). A Least Squares Correction for Selectivity Bias. *Econometrica*, **48**, 1815-1820.

Rubin D. B. (1976). *Inference and Missing Data. Biometrika*, **63**, 581-590.

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.

National Center for Health Statistics (1992). Third National Health and Nutrition Examination Survey. *Vital and Health Statistics Series 2*, **113**.

Park, T. (1998). An Approach to Categorical Data Nonignorable Nonresponse. *Biometrics*, **54**, 1579-1590.

Park, T. and Brown, M. B. (1994). Models for Categorical Data with Nonignorable Nonresponse. *Journal of the American Statistical Association*, **89**, 44-52.

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.

Stasny, E. A., Kadane, J. B., and Fritsch, K. S. (1991). On the Fairness of death Penalty Jurors: A Comparison of Bayesian Models with Diffrent levels of Hierarchy and Various Missing Data Mechanisms. *Journal of the American Statistical Association*, **93**, 464-477.