Statistica Sinica **28** (2018), 1839-1866
doi:https://doi.org/10.5705/ss.202016.0121

BAYESIAN SMALL AREA MODELS FOR THREE-WAY CONTINGENCY TABLES WITH NONIGNORABILITY

Namkyo Woo, Balgobin Nandram and Dalho Kim

Korea Appraisal Board, Worcester Polytechnic Institute and Kyungpook National University

Abstract: We show how to use Bayesian uncertainty analysis to study several three-way contingency tables, each obtained from a single area, when one, two or three categories are missing. This is an extension of Nandram and Woo (2015) to cover small areas. One approach to analyze these data is to construct several tables (one complete and the others incomplete) with each table corresponding to one or more missing categories. When tables are incomplete and nonignorable nonresponse models are used, there are nonidentifiable parameters. To deal with these parameters, we describe five hierarchical Bayesian models, which are an ignorable nonresponse model and four nonignorable nonresponse models. Rather than performing a sensitivity analysis, we perform the Bayesian uncertainty analysis by placing priors on the nonidentifiable parameters. This is done to reduce the effects of the nonidentifiable parameters that is accomplished by projecting the parameters to a lower dimensional space and allowing the reduced set of parameters to share a common distribution. Also, this procedure allows a "borrowing of strength" from larger areas to improve estimation in smaller areas. We use the griddy Gibbs sampler to fit our models and we use goodness-of-fit procedures to assess model fit. We use an illustrative example and a simulation study to compare our models when inference is made about finite population proportions of the cells of the three-way tables.

Key words and phrases: Bayesian uncertainty analysis, griddy Gibbs sampler, model diagnostics, nonidentifiable parameters, nonignorable nonresponse model.

1. Introduction

In survey sampling, data may consist of contingency tables with missing cells. We consider the problem of nonignorable nonresponse for three-way ($r \times c \times u$) categorical tables, each obtained from a single small area. There are both item and unit nonresponses. Unit nonresponse occurs when none of the categorical variables is observed. Item nonresponse occurs when a categorical variable is not observed. For both item and unit nonresponse, none of the possible categories is observed for any of the missing categorical variables. We do not know how different are the observed data and the missing data. In fact, when there are some differences between observed and missing data, it is better to use a nonignorable missing data model because of its generality.

It is pertinent to distinguish between ignorable and nonignorable nonresponse models. These are associated with the missing data mechanism (Little and Rubin (2002)), and there are three types of missing data mechanism. Missing completely at random (MCAR) occurs if the missingness is independent of both the observed and the unobserved data, and missing at random (MAR) when conditional on the observed data, missingness is independent of the unobserved data. Missing not at random (MNAR) is neither MCAR nor MAR. While under the MCAR mechanism, missing data do not contribute to the analysis, under the MAR mechanism incomplete data may be relevant. Models for MCAR and MAR are called ignorable, and models for MNAR missing data mechanism are called nonignorable. Since the missing data are different from the observed data, the main issue of MNAR is how to fill in nonresponse. However, the general difficulty with nonignorable nonresponse models is that there are nonidentifiable parameters (e.g., see Nandram and Choi (2010)). Nandram and Woo (2015), henceforth NW, described Bayesian uncertainty analysis for nonignorable nonresponse models that incorporate a sensitivity analysis directly into the Bayesian nonignorable nonresponse models. This methodology was done for a single area.

There are both item and unit nonresponses in a three-way contingency table. To deal with missing data, we consider the data to consist of eight tables. One table is for complete data and seven tables are for incomplete data - one table for missing row, one table for missing column, one table for missing length, one table for both missing row and column, one table for both missing row and length, one table for both missing column and length, and a table for which neither row, column nor length has been observed. We model the observed and missing data from these eight separate tables. We can fit a multinomial data model to these tables including the nonresponse data. Pioneering work was done by Chen and Fienberg (1974) who provided a non-Bayesian analysis for incomplete two-way tables. Another approach uses log-linear model for nonignorable nonresponse (e.g., Draper (1995); Barker, Rosenberger and DerSimonian (1992); Rubin, Stern and Behovar (1995)).

Nandram and Choi (2002) used an expansion model to study nonignorable nonresponse binary data and Nandram and Katzoff (2012) used a similar model for polychotomous data. If a centering parameter is set to unity, the expansion model of nonignorable nonresponse degenerates into an ignorable nonresponse model. This can be used to express uncertainty about ignorability (Forster and Smith (1998)). We use an idea of Nandram, Cox and Choi (2005) and Nandram et al. (2005) who assumed an ignorable model, obtained samples of the response probabilities, and used these sampled response probabilities to fit the response probabilities of a nonignorable nonresponse model while controlling its parameters.

If we have no information about the missing data, we do not know how to fill in the missing cells. We need to deal with subjectivity and imprecision which are, respectively, due to missingness and sampling (Molenberghs, Kenward and Goetghebeur (2001)). By fitting several plausible overspecified models, it provides an expression of uncertainty about the parameters of interest. Molenberghs et al. (1999) gave some examples of categorical tables in which different nonignorable nonresponse models have the same fit to the observed data but the prediction of the missing counts is different. That is, nonignorable nonresponse models cannot be examined using the observed data alone even if they fit well, the plausibility of the model assumptions needs to be examined carefully.

This paper represents an extension of NW to accommodate small areas, and it has four more sections. In Section 2, we give a review of NW. In Section 3, we describe competing hierarchical Bayesian nonignorable nonresponse models of small areas using Bayesian uncertainty analysis. In Section 4, we illustrate our methodology with public-use data from ten states in the third National Health and Nutrition Examination Survey (NHANES III), and we report on a simulation study to assess our methodology. In Section 5, we have concluding remarks.

2. A Review of Nandram and Woo (2015)

In a nonignorable nonresponse model, a sensitivity analysis is necessary to study the effects of nonidentifiable parameters on the parameters of interest. Typically, the sensitivity analysis is performed by setting the nonidentifiable parameters at various plausible values. Instead of the sensitivity analysis for the nonidentifiable parameters, we extend Nandram and Woo (2015), NW, to accommodate small areas. Therefore, it is pertinent to review Bayesian uncertainty analysis in NW.

In the frequentist point of view, Molenberghs, Kenward and Goetghebeur (2001) provided a general principle for missing data. The parameter space consists of two sets (η, ν), where η is a minimal set of parameters which can be estimated when ν is specified. Here, η is called the estimable parameter and ν

the sensitivity parameter. Since the value of ν is specified, we have an estimate $\hat{\eta}(\nu)$. The range of these estimates over all plausible values of ν is the interval of ignorance. The union of the $100(1-\alpha)\%$ confidence intervals over all plausible values of ν gives the uncertainty interval.

NW formulated the work of Molenberghs, Kenward and Goetghebeur (2001) within the Bayesian framework. This permitted a 'Bayesian uncertainty interval' for the finite population parameters to include both subjectivity and imprecision. NW considered two sets of parameters and put prior distributions on them. Thus,

$$p(\eta,\nu) = p(\eta|\nu)p(\nu),$$

where again η is the set of parameters of interest and ν is the set of nonidentifiable parameters. The prior on ν is specified on a set of plausible values of ν. If ν is specified as in a sensitivity analysis, η will be identified. Thus, one way to specify $p(\nu)$ is to put a uniform distribution over all plausible values of ν. Rather than performing a sensitivity analysis, a Bayesian uncertainty analysis puts a prior on the nonidentifiable parameters. This is related to analysis of biases in observational studies (Greenland (2009)). The uncertainty interval is the $(1-\alpha)$ credible interval for η when a prior distribution is placed on the nonidentifiable parameters.

For a Bayesian uncertainty analysis there are two strategies, projection and pooling in a three-way categorical table. In the projection strategy, we can project parameters to a lower dimensional space. NW considered a two-way (2×2) table. Let the count in $(1,1)$ cell be z, the first row total be x, and the first column total be y, the corresponding cell probabilities are θ, p and q. Then for a random sample of n, NW used the joint density

$$p(x,y,z|\theta,p,q) = \frac{n!\theta^z(p-\theta)^{(x-z)}(q-\theta)^{(y-z)}(1-p-q+\theta)^{(n-x-y+z)}}{z!(x-z)!(y-z)!(n-x-y+z)!},$$

where $0 < z \le x, y < n$ and $0 < \theta < p, q < 1$.

NW considered a three-way $(2\times 2\times 2)$ table. They can generalize this multinomial distribution and they assumed each level of the first category with proportion of r and count of w gives rise to a 2×2 table. Let the counts in the two tables be (x_1,y_1,z_1) and (x_2,y_2,z_2). Then NW used the joint probability mass function of $(w,x_1,y_1,z_1,x_2,y_2,z_2)$

$$\begin{aligned} &p(w,x_1,y_1,z_1,x_2,y_2,z_2|r,\theta_1,p_1,q_1,\theta_2,p_2,q_2) \\ &= \frac{n!}{w!(n-w)!} r^w (1-r)^{(n-w)} \end{aligned} \tag{2.1}$$

$$\times \frac{w!\theta_1^{z_1}(p_1-\theta_1)^{(x_1-z_1)}(q_1-\theta_1)^{(y_1-z_1)}(1-p_1-q_1+\theta_1)^{(w-x_1-y_1+z_1)}}{z_1!(x_1-z_1)!(y_1-z_1)!(w-x_1-y_1+z_1)!}$$
$$\times \frac{(n-w)!\theta_2^{z_2}(p_2-\theta_2)^{(x_2-z_2)}(q_2-\theta_2)^{(y_2-z_2)}(1-p_2-q_2+\theta_2)^{(n-w-x_2-y_2+z_2)}}{z_2!(x_2-z_2)!(y_2-z_2)!(n-w-x_2-y_2+z_2)!},$$

where $0 < z_1 \le x_1, y_1 < w$, $0 < z_2 \le x_2, y_2 < n-w$, $0 < r < 1$, $0 < \theta_1 < p_1, q_1 < 1$, $0 < \theta_2 < p_2, q_2 < 1$.

The eight tables have distinct parameters $(r_t, \theta_{t1}, q_{t1}, q_{t1}, \theta_{t2}, q_{t2}, q_{t2})$ with the restrictions shown below (2.1), $t = 1, \ldots, 8$, with probability mass functions similar to (2.1). The number of observations in the t^{th} and table is n_t with $\sum_1^8 n_t = n$. Except for the complete table, these parameters are nonidentifiable. Therefore, the nonignorable nonresponse model has a joint probability mass function of $(w, x_1, y_1, z_1, x_2, y_2, z_2)$ given $(r_t, \theta_{t1}, q_{t1}, q_{t1}, \theta_{t2}, q_{t2}, q_{t2})$ are independent with probability mass functions similar to (2.1).

In the pooling strategy, NW allowed the reduced set of parameters to share a common distribution. They specified a joint conjugate prior distribution for $(r_t, \theta_{t1}, q_{t1}, q_{t1}, \theta_{t2}, q_{t2}, q_{t2})$ as

$$\begin{aligned}
& r_t|\mu_0, \tau \overset{\text{i.i.d.}}{\sim} \text{Beta}\{\mu_0\tau, (1-\mu_0)\tau\}, \\
& (\theta_{tj}, p_{tj}-\theta_{tj}, q_{tj}-\theta_{tj}, 1-p_{tj}-q_{tj}+\theta_{tj})|\mu_1, \mu_2, \mu_3, \tau \overset{\text{i.i.d.}}{\sim} \\
& \text{Dirichlet}\{\mu_1\tau, (\mu_2-\mu_1)\tau, (\mu_3-\mu_1)\tau, (1-\mu_2-\mu_3+\mu_1)\tau\},
\end{aligned}$$

where $j = 1, 2$, $t = 1, \ldots, 8$, and $0 < \mu_0 < 1, 0 < \mu_1 < \mu_2, \mu_3 < 1$, $\tau > 0$. If the parameters $\mu_0, \mu_1, \mu_2, \mu_3$, and τ are specified, then the model is well identified. For sensitivity analysis we can take various values of these parameters, but it is more sensible to perform a Bayesian uncertainty analysis. NW treated these parameters as hyper parameters, and placed priors on them. This permits a study of subjectivity to provide a coherent method to obtain an uncertainty interval for the finite population proportion. It does not work completely; NW needed some adjustment that puts a bound on μ_1, B, and assumed a shrinkage prior for τ. Therefore, they assumed that hyper parameters are constrained on the set $S = \{(\mu_0, \mu_1, \mu_2, \mu_3) : 0 < \mu_0 < 1, 0 < B < \mu_1 < \mu_2, \mu_3 < 1\}$ and

$$\begin{aligned}
& (\mu_0, \mu_1, \mu_2, \mu_3)|B \sim \text{Uniform}(S), \\
& B \sim \text{Uniform}(a, b), \quad p(\tau) = \frac{1}{(1+\tau)^2}, \quad \tau > 0.
\end{aligned}$$

This passes on the nonidentifiable effects to a smaller set of hyper parameters. Thus, projection and pooling lead to a reduced set of nonidentifiable parameters.

In this paper, we extend NW to accommodate small areas. We provide a

Bayesian uncertainty analysis for nonignorable nonresponse in three-way contingency tables obtained from small areas. In small area estimation, it is a standard practice to assume that the area effects are exchangeable. This assumption is accommodated by allowing the area effects to have a common parametric distribution, and there is a "borrowing of strength" of the data from larger areas to improve the reliability in the estimates of the model parameters corresponding to the smaller areas.

3. The Nonignorable Nonresponse Models of Small Areas

In this paper, we index small areas by $i = 1, \ldots, A$; rows by $j = 1, \ldots, r$; columns by $k = 1, \ldots, c$; lengths by $\ell = 1, \ldots, u$; and the eight tables by $t = 1, \ldots, T, T = 8$. For a three-way categorical table, let $J_{its} = 1$ if the s^{th} and individual in i^{th} and area belongs to the t^{th} and table and $J_{its} = 0$ for the other seven tables, and let $I_{ijk\ell s} = 1$ if the s^{th} and individual in i^{th} and area belongs to the cell (j, k, ℓ) of the three-way table and $I_{ijk\ell s} = 0$ for all other cells. Also let $w_{itjk\ell s} = J_{its} I_{ijk\ell s}$. Now, let $p_{itjk\ell}$ be the probability that an individual belongs to cell (j, k, ℓ) of t^{th} and sub-table in the three-way table for i^{th} and area, and let π_{it} be the probability that an individual belongs to the t^{th} and sub-table for i^{th} and area.

In the ignorable nonresponse model, parameters $p_{itjk\ell}$ do not depend on t, and the parameters are identifiable and estimable. Yet the ignorable nonresponse model contains imprecision and subjectivity of the relationship between respondents and nonrespondents, subjectivity and imprecision cannot be separated. As pointed out by a referee, for categorical variable j, we can write

$$P(j \text{ missing}|j = 1) = \frac{\sum_{t \in J(t)} \sum_{k=1}^{c} \sum_{\ell=1}^{u} p_{it1k\ell} \pi_{it}}{\sum_{t \in T} \sum_{k=1}^{c} \sum_{\ell=1}^{u} p_{it1k\ell} \pi_{it}}, \tag{3.1}$$

where $J(t)$ is the set of tables in which categorical variable j is not observed, and T is the set of all tables. Clearly, if the $p_{itjk\ell}$ depend on t, (3.1) is not necessarily equal to $P(j \text{ missing}|j = 2)$. For the ignorable nonresponse model with $p_{itjk\ell} = p_{ijk\ell}$, the $\sum_{k=1}^{c} \sum_{\ell=1}^{u} p_{i1k\ell}$ in the numerator cancels with that in the denominator so that $P(j \text{ missing}|j = 1) = P(j \text{ missing}|j = 2)$.

The ignorable nonresponse model arises from a MAR mechanism but obviously the model can be incorrect and a more general model may include the ignorable nonresponse model as a special case. However, when there may be information in the nonresponse data, we consider the nonignorable nonresponse model which includes a possible difference between observed and missing data (i.e., subjectivity). In the nonignorable nonresponse model, the parameters π_{it} are identifiable, but the parameters $p_{ijk\ell t}$ are not identifiable for $t = 2, \ldots, 8$.

Under simple random sampling, our basic model is

$$\boldsymbol{J}_{is}|\boldsymbol{\pi}_i \overset{\text{i.i.d.}}{\sim} \text{Multinomial}(1, \boldsymbol{\pi}_i),$$
$$\boldsymbol{I}_{is}|J_{its} = 1, \boldsymbol{p}_{it} \overset{\text{i.i.d.}}{\sim} \text{Multinomial}(1, \boldsymbol{p}_{it}),$$

where $\boldsymbol{J}_{is} = (J_{i1s}, \ldots, J_{iTs})'$, $\boldsymbol{I}_{is} = (I_{i111s,\ldots,I_{ircus}})'$, $\boldsymbol{\pi}_i = (\pi_{i1}, \ldots, \pi_{iT})'$, $\boldsymbol{p}_{it} = (p_{it111}, \ldots, p_{itrcu})'$.

Let $\psi_{ijk\ell t} = \pi_{it} p_{itjk\ell}$. Then because $\sum_t \pi_{it} = 1$ and $\sum_{jk\ell} p_{itjk\ell} = 1$ for each $t = 1, \ldots, 8$, $\sum_t \sum_{jk\ell} \pi_{it} p_{itjk\ell} = 1$. Letting $w_{itjk\ell s} = J_{its} I_{ijk\ell s}$, it follows that

$$\boldsymbol{w}_{is}|\boldsymbol{p}_i, \boldsymbol{\pi}_i \overset{\text{i.i.d.}}{\sim} \text{Multinomial}(1, \boldsymbol{\psi}_i),$$

where $\boldsymbol{w}_{is} = (w_{i1111s}, \ldots, w_{iTrcus})'$, $\boldsymbol{p}_i = (p_{i1111}, \ldots, p_{iTrcu})'$, $\boldsymbol{\psi}_i = (\psi_{i1111}, \ldots, \psi_{iTrcu})'$.

While the parameters π_{it} are identifiable, the parameters $p_{itjk\ell}$ are not identifiable for $t = 2, \ldots, 8$. Given the constraints and the observed data, inference for the $p_{itjk\ell}$ is independent of the π_{it}. To reduce the effects of nonidentifiable parameters, we consider the nonignorable nonresponse model using Bayesian uncertainty analysis as in NW.

In the first strategy of Bayesian uncertainty analysis, we can project $p_{itjk\ell}$ to a lower dimensional space. This can be done by expressing the $p_{itjk\ell}$ as functions of a reduced set of parameters. In our current work, we have n_{it} individuals in t^{th} and table for i^{th} and area and $n_i = \sum_{t=1}^{8} n_{it}$. Also, for the eight tables the cell counts are $z_{itj} = \sum_{s=1}^{n_{it}} w_{itj11s}$, $x_{itj} = \sum_{s=1}^{n_{it}} \sum_{\ell=1}^{u} w_{itj1\ell s}$, $y_{itj} = \sum_{s=1}^{n_{it}} \sum_{k=1}^{c} w_{itjk1s}$, $w_{itj} = \sum_{s=1}^{n_{it}} \sum_{k=1}^{c} \sum_{\ell=1}^{u} w_{itjk\ell s}$, and the corresponding superpopulation proportions are θ_{itj}, p_{itj}, q_{itj} and r_{it}. We describe the data structure in Figure 1 and Table 1.

The joint probability mass function under the nonignorable nonresponse model is

$$\begin{aligned}
&p(w_{it1}, x_{it1}, y_{it1}, z_{it1}, x_{it2}, y_{it2}, z_{it2}|r_{it}, \theta_{it1}, p_{it1}, q_{it1}, \theta_{it2}, p_{it2}, q_{it2}) \\
&= \frac{n_{it}!}{w_{it1}! w_{it2}!} r_{it}^{w_{it1}} (1 - r_{it})^{w_{it2}} \times \\
&\prod_{j=1}^{2} \frac{w_{itj}! \theta_{itj}^{z_{itj}} (p_{itj} - \theta_{itj})^{x_{itj} - z_{itj}} (q_{itj} - \theta_{itj})^{y_{itj} - z_{itj}} (1 - p_{itj} - q_{itj} + \theta_{itj})^{w_{itj} - x_{itj} - y_{itj} + z_{itj}}}{z_{itj}! (x_{itj} - z_{itj})! (y_{itj} - z_{itj})! (w_{itj} - x_{itj} - y_{itj} + z_{itj})!},
\end{aligned}$$

where $w_{it2} = n_{it} - w_{it1}$.

In the next strategy, we can allow the reduced set of parameters to share a

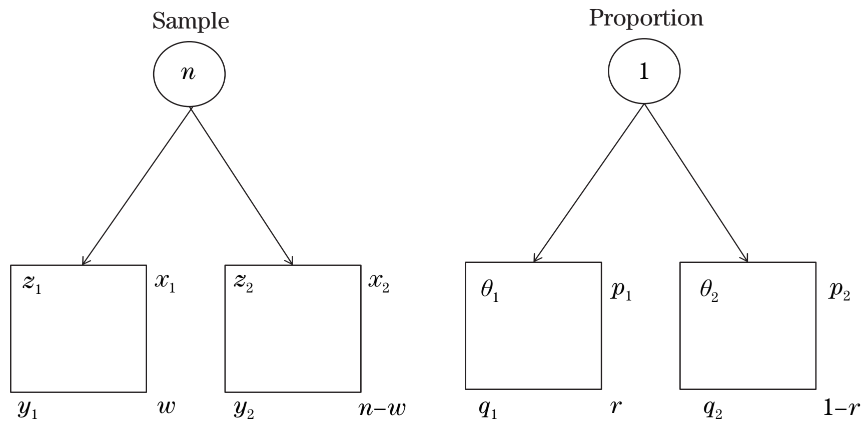

Figure 1. The structure of proportions and data for three-way categorical table.

Table 1. Tables of observed and missing counts for a general table.

Table	Observed	Missing
t1	$x_k, y_k, z_k, (k = 1, 2), w$	None
t2	$x_1 + x_2, y_1 + y_2, z_1 + z_2$	x_1, y_1, z_1, w
t3	y_1, y_2, w	x_1, x_2, z_1, z_2
t4	x_1, x_2, w	y_1, y_2, z_1, z_2
t5	w	$x_1, x_2, y_1, y_2, z_1, z_2$
t6	$x_1 + x_2$	$x_1, x_2, y_1, y_2, z_1, z_2, w$
t7	$y_1 + y_2$	$x_1, x_2, y_1, y_2, z_1, z_2, w$
t8	None	$x_k, y_k, z_k, (k = 1, 2), w$

NOTE : Table t1 is complete; each of Tables t2, t3, t4 has one category missing; each of Tables t5, t6, t7 has two categories missing, and Table t8 has all categories missing.

common distribution. This passes on the nonidentifiable effects to a smaller set of hyper parameters. In our model, we use a Beta prior density and a Dirichlet prior density for categorical cell probabilities,

$$r_{it}|\mu_0, \tau \overset{\text{i.i.d.}}{\sim} \text{Beta}\{\mu_0\tau, (1-\mu_0)\tau\},$$

$$(\theta_{itj}, p_{itj} - \theta_{itj}, q_{itj} - \theta_{itj}, 1 - p_{itj} - q_{itj} + \theta_{itj})|\mu_1, \mu_2, \mu_3, \tau \overset{\text{i.i.d.}}{\sim}$$
$$\text{Dirichlet}\{\mu_1\tau, (\mu_2 - \mu_1)\tau, (\mu_3 - \mu_1)\tau, (1 - \mu_2 - \mu_3 + \mu_1)\tau\},$$

where $0 < \mu_0 < 1, 0 < \mu_1 < \mu_2, \mu_3 < 1, \tau > 0$.

These two strategies lead to a reduced set of nonidentifiable parameters. The specification of priors on these parameters is the Bayesian uncertainty analysis. We consider three ways to specify priors. Henceforth, we focus on the $2 \times 2 \times 2$ table. The methodology for a general $r \times c \times u$ table is similar, but the notations are more complex.

To lay our foundation, we first describe the ignorable nonresponse model (Ig) in which the parameters $r_{it}, \theta_{itj}, p_{itj}$, and q_{itj} do not depend on t, assuming $\theta_{itj} = \theta_i$, $p_{itj} = p_i$, and $q_{itj} = q_i$, $j = 1, 2$. Therefore, the joint probability mass function of the ignorable nonresponse model is

$$
\begin{aligned}
&p(w_{i1}, x_{i1}, y_{i1}, z_{i1}, x_{i2}, y_{i2}, z_{i2} | r_i, \theta_i, p_i, q_i) \\
&= \frac{n_i!}{w_{i1}!(n_i - w_{i1})!} r_i^{w_{i1}} (1 - r_i)^{n_i - w_{i1}} \\
&\quad \times \frac{w_{i1}! \theta_i^{z_{i1}+z_{i2}} (p_i - \theta_i)^{x_{i1}+x_{i2}-z_{i1}-z_{i2}} (q_i - \theta_i)^{y_{i1}+y_{i2}-z_{i1}-z_{i2}}}{z_{i1}! z_{i2}! (x_{i1} - z_{i1})! (x_{i2} - z_{i2})! (y_{i1} - z_{i1})! (y_{i2} - z_{i2})!} \\
&\quad \times \frac{(1 - p_i - q_i + \theta_i)^{n_i - x_{i1} - x_{i2} - y_{i1} - y_{i1} + z_{i1} + z_{i2}}}{(w_{i1} - x_{i1} - y_{i1} + z_{i1})! (n_i - w_{i1} - x_{i2} - y_{i2} + z_{i2})!}.
\end{aligned}
$$

We take priors for parameters in the ignorable nonresponse model as

$$
\begin{aligned}
&r_i | \mu_0, \tau \overset{\text{i.i.d.}}{\sim} \text{Beta}\{\mu_0 \tau, (1 - \mu_0)\tau\}, \\
&(\theta_i, p_i - \theta_i, q_i - \theta_i, 1 - p_i - q_i + \theta_i) | \mu_1, \mu_2, \mu_3, \tau \overset{\text{i.i.d.}}{\sim} \\
&\text{Dirichlet}\{\mu_1 \tau, (\mu_2 - \mu_1)\tau, (\mu_3 - \mu_1)\tau, (1 - \mu_2 - \mu_3 + \mu_1)\tau\}, \\
&(\mu_0, \mu_1, \mu_2, \mu_3) \sim \text{Uniform}(S), \\
&p(\tau) = \frac{1}{(1 + \tau)^2}, \quad \tau > 0,
\end{aligned}
$$

where hyper parameters are constrained on the set $S = \{(\mu_0, \mu_1, \mu_2, \mu_3) : 0 < \mu_0 < 1, 0 < \mu_1 < \mu_2, \mu_3 < 1\}$.

3.1. Nonignorable nonresponse models

In this section we describe four nonignorable nonresponse models that extend the ignorable nonresponse model. The four models differ in their prior specifications.

First, we describe the nonignorable nonresponse model (Nig1) with data-based priors that are needed because of the nonidentifiable parameters in Nig1. We assume that $\theta_{itj} = \theta_{it}$, $p_{itj} = p_{it}$ and $q_{itj} = q_{it}$, $j = 1, 2$. The joint probability mass function of the nonignorable nonresponse model (Nig1, Nig2) is

$$
\begin{aligned}
&p(w_{it1}, x_{it1}, y_{it1}, z_{it1}, x_{it2}, y_{it2}, z_{it2} | r_{it}, \theta_{it}, p_{it}, q_{it}) \\
&= \frac{n_{it}!}{w_{it1}!(n_{it} - w_{it1})!} r_{it}^{w_{it1}} (1 - r_{it})^{n_{it} - w_{it1}} \\
&\quad \times \frac{w_{it1}! \theta_{it}^{z_{it1}+z_{it2}} (p_{it} - \theta_{it})^{x_{it1}+x_{it2}-z_{it1}-z_{it2}} (q_{it} - \theta_{it})^{y_{it1}+y_{it2}-z_{it1}-z_{it2}}}{z_{it1}! z_{it2}! (x_{it1} - z_{it1})! (x_{it2} - z_{it2})! (y_{it1} - z_{it1})! (y_{it2} - z_{it2})!}
\end{aligned}
$$

$$\times \frac{(1-p_{it}-q_{it}+\theta_{it})^{n_{it}-x_{it1}-x_{it2}-y_{it1}-y_{it2}+z_{it1}+z_{it2}}}{(w_{it1}-x_{it1}-y_{it1}+z_{it1})!(n_{it}-w_{it1}-x_{it2}-y_{it2}+z_{it2})!}.$$

We take priors for parameters as

$$r_{it}|\mu_0, \tau \overset{\text{i.i.d.}}{\sim} \text{Beta}\{\mu_0\tau, (1-\mu_0)\tau\}, \tag{3.2}$$

$$(\theta_{it}, p_{it}-\theta_{it}, q_{it}-\theta_{it}, 1-p_{it}-q_{it}+\theta_{it})|\mu_1, \mu_2, \mu_3, \tau \overset{\text{i.i.d.}}{\sim} \tag{3.3}$$
$$\text{Dirichlet}\{\mu_1\tau, (\mu_2-\mu_1)\tau, (\mu_3-\mu_1)\tau, (1-\mu_2-\mu_3+\mu_1)\tau\},$$

where $0 < \mu_0 < 1, 0 < \mu_1 < \mu_2, \mu_3 < 1, \tau > 0$.

In Nig1 we use a data-based prior for $\mu_0, \mu_1, \mu_2, \mu_3$, and τ, considering priors for them based on data that are recorded when we fit the ignorable nonresponse model. When we fit the ignorable nonresponse model, after a burn-in period, we select the H iterates, $(\mu_0^{(h)}, \mu_1^{(h)}, \mu_2^{(h)}, \mu_3^{(h)})$, $h = 1, \ldots, H$, of $\mu_0, \mu_1, \mu_2, \mu_3, \tau$. In addition, we need to estimate the hyper parameters, $\mu_{00}, \mu_{10}, \mu_{20}, \mu_{30}$ and α_0, β_0 of priors. To do so, we assume that these H iterates follow the model,

$$\mu_0^{(h)}|\mu_{00}, \tau_0 \overset{\text{i.i.d.}}{\sim} \text{Beta}\{\mu_{00}\tau_0, (1-\mu_{00})\tau_0\}, \quad 0 < \mu_{00} < 1, \tau_0 > 0,$$
$$(\mu_1^{(h)}, \mu_2^{(h)}-\mu_1^{(h)}, \mu_3^{(h)}-\mu_1^{(h)}-\mu_2^{(h)}-\mu_3^{(h)}+\mu_1^{(h)})|\mu_{10}, \mu_{20}, \mu_{30}, \tau_0 \overset{\text{i.i.d.}}{\sim}$$
$$\text{Dirichlet}\{\mu_{10}\tau_0, (\mu_{20}-\mu_{10})\tau_0, (\mu_{30}-\mu_{10})\tau_0, (1-\mu_{20}-\mu_{30}+\mu_{10})\tau_0\},$$
$$\tau^{(h)} \overset{\text{i.i.d.}}{\sim} \text{Gamma}(\alpha_0, \beta_0),$$

where $0 < \mu_{10} < \mu_{20}, \mu_{30} < 1$. The posterior means of $\mu_{00}, \mu_{10}, \mu_{20}, \mu_{30}$, τ_0 are $\mu_{00}^{(0)}, \mu_{10}^{(0)}, \mu_{20}^{(0)}, \mu_{30}^{(0)}$, $\tau_0^{(0)}$ and estimates of α_0, β_0 are obtained as $\alpha_0^{(0)}, \beta_0^{(0)}$ (maximum likelihood estimates). Now the data-based prior distributions on $(\mu_0, \mu_1, \mu_2, \mu_3, \tau)$ are

$$\mu_0|\mu_{00}^{(0)}, \tau_0^{(0)} \sim \text{Beta}\{\mu_{00}^{(0)}\tau_0^{(0)}, (1-\mu_{00}^{(0)})\tau_0^{(0)}\},$$
$$(\mu_1, \mu_2-\mu_1, \mu_3-\mu_1, 1-\mu_2-\mu_3+\mu_1)|\mu_{10}^{(0)}, \mu_{20}^{(0)}, \mu_{30}^{(0)}, \tau_0^{(0)} \sim$$
$$\text{Dirichlet}\{\mu_{10}^{(0)}\tau_0^{(0)}, (\mu_{20}^{(0)}-\mu_{10}^{(0)})\tau_0^{(0)}, (\mu_{30}^{(0)}-\mu_{10}^{(0)})\tau_0^{(0)}, (1-\mu_{20}^{(0)}-\mu_{30}^{(0)}+\mu_{10}^{(0)})\tau_0^{(0)}\},$$
$$\tau \sim \text{Gamma}(\alpha_0^{(0)}, \beta_0^{(0)}).$$

Second, we keep the assumption $\theta_{itj} = \theta_{it}$, $p_{itj} = p_{it}$, and $q_{itj} = q_{it}$, $j = 1, 2$ and we consider the nonignorable nonresponse model (Nig2) without data-based priors; we take priors for parameters as (3.2)–(3.3). The hyper parameters $\mu_0, \mu_1, \mu_2, \mu_3$ are constrained on the set S. If $\mu_0, \mu_1, \mu_2, \mu_3$, and τ are specified, then the model is well identified. One way to do the adjustment is to put a bound on μ_1, say B, and take a proper diffuse prior for τ. Therefore, these parameters are constrained on the set $S = \{(\mu_0, \mu_1, \mu_2, \mu_3) : 0 < \mu_0 < 1, 0 < B \leq \mu_1 <$

$\mu_2, \mu_3 < 1\}$; see NW. Then we take priors for hyper parameters as

$$(\mu_0, \mu_1, \mu_2, \mu_3)|B \sim \text{Uniform}(S), \tag{3.4}$$

$$B \sim \text{Uniform}(a, b), \tag{3.5}$$

$$p(\tau) = \frac{1}{(1+\tau)^2}, \quad \tau > 0, \tag{3.6}$$

where a and b should be specified. For example, we can give a nonparametric interval of (a, b). A lower bound a might assume that only observed data are included in cell $(1,1,1)$ and an upper bound b might assume that all missing data are included in cell $(1,1,1)$, thereby forming a pessimistic-optimistic range for B.

In Ig, Nig1, and Nig2, we assume the cell probabilities $\theta_{itj} = \theta_{it}$, $p_{itj} = p_{it}$, and $q_{itj} = q_{it}$, $j = 1, 2$. Next, we distinguish these cell probabilities in the nonignorable nonresponse model. Nig3 is a general model, an extension of Nig1 that uses a data-based prior, and Nig4 is a general model, an extension of Nig2 that does not use data-based prior. Letting $w_{it2} = n_{it} - w_{it1}$, the joint probability mass function of the nonignorable nonresponse model (Nig3, Nig4) is

$$\begin{aligned}
&p(w_{it1}, x_{it1}, y_{it1}, z_{it1}, x_{it2}, y_{it2}, z_{it2}|r_{it}, \theta_{it1}, p_{it1}, q_{it1}, \theta_{it2}, p_{it2}, q_{it2}) \\
&= \frac{n_{it}!}{w_{it1}!(n_{it} - w_{it1})!} r_{it}^{w_{it1}} (1 - r_{it})^{n_{it} - w_{it1}} \times \\
&\prod_{j=1}^{2} \frac{w_{itj}! \theta_{itj}^{z_{itj}} (p_{itj} - \theta_{itj})^{x_{itj} - z_{itj}} (q_{itj} - \theta_{itj})^{y_{itj} - z_{itj}} (1 - p_{itj} - q_{itj} + \theta_{itj})^{w_{itj} - x_{itj} - y_{itj} + z_{itj}}}{z_{itj}!(x_{itj} - z_{itj})!(y_{itj} - z_{itj})!(w_{itj} - x_{itj} - y_{itj} + z_{itj})!}.
\end{aligned}$$

We take priors for parameters as

$$\begin{aligned}
&r_{it}|\mu_0, \tau \overset{\text{i.i.d.}}{\sim} \text{Beta}\{\mu_0\tau, (1-\mu_0)\tau\}, \\
&(\theta_{itj}, p_{itj} - \theta_{itj}, q_{itj} - \theta_{itj}, 1 - p_{itj} - q_{itj} + \theta_{itj})|\mu_1, \mu_2, \mu_3, \tau \overset{\text{i.i.d.}}{\sim} \\
&\text{Dirichlet}\{\mu_1\tau, (\mu_2 - \mu_1)\tau, (\mu_3 - \mu_1)\tau, (1 - \mu_2 - \mu_3 + \mu_1)\tau\}.
\end{aligned}$$

For the hyper parameters $\mu_0, \mu_1, \mu_2, \mu_3$, and τ in Nig3, we take data-based priors similar to those in Nig1, and for the hyper parameters in Nig4, we take priors similar to Nig2 (without data-based priors; see (3.4)–(3.6)).

Let $r_{it1} = r_{it}, r_{it2} = 1 - r_{it}$ and $\boldsymbol{d}_{mis}$ and $\boldsymbol{d}_{obs}$ denote all missing data and all observed data, respectively. Then, the joint posterior density of all parameters and missing values is

$$\pi(\boldsymbol{r}, \boldsymbol{\theta}, \boldsymbol{p}, \boldsymbol{q}, \boldsymbol{d}_{mis}, \mu_0, \mu_1, \mu_2, \mu_3, \tau|\boldsymbol{d}_{obs})$$

$$
\begin{aligned}
&\propto \prod_{i=1}^{A}\prod_{t=1}^{T}\prod_{j=1}^{2}\Bigg[\frac{(r_{itj}\theta_{itj})^{z_{itj}}}{z_{itj}!}\frac{\{r_{itj}(p_{itj}-\theta_{itj})\}^{(x_{itj}-z_{itj})}}{(x_{itj}-z_{itj})!}\frac{\{r_{itj}(q_{itj}-\theta_{itj})\}^{(y_{itj}-z_{itj})}}{(y_{itj}-z_{itj})!}\\
&\quad\times\frac{\{r_{itj}(1-p_{itj}-q_{itj}+\theta_{itj})\}^{(w_{itj}-x_{itj}-y_{itj}+z_{itj})}}{(w_{itj}-x_{itj}-y_{itj}+z_{itj})!}\\
&\quad\times\frac{\theta_{itj}^{\mu_1\tau-1}}{\Gamma(\mu_1\tau)}\frac{(p_{itj}-\theta_{itj})^{(\mu_2-\mu_1)\tau-1}}{\Gamma((\mu_2-\mu_1)\tau)}\frac{(q_{itj}-\theta_{itj})^{(\mu_3-\mu_1)\tau-1}}{\Gamma((\mu_3-\mu_1)\tau)}\\
&\quad\times\frac{(1-p_{itj}-q_{itj}+\theta_{itj})^{(1-\mu_2-\mu_3+\mu_1)\tau-1}}{\Gamma((1-\mu_2-\mu_3+\mu_1)\tau)}\Bigg]\\
&\quad\times\prod_{i=1}^{A}\prod_{t=1}^{T}\left\{\frac{r_{it}^{\mu_0\tau-1}}{\Gamma(\mu_0\tau)}\frac{(1-r_{it})^{(1-\mu_0)\tau-1}}{\Gamma((1-\mu_0)\tau)}\right\}\times\frac{\{\Gamma(\tau)\}^{3AT}}{(1+\tau)^2},
\end{aligned}
$$

where $\boldsymbol{r} = (r_{111},\ldots,r_{AT2})'$, $\boldsymbol{\theta} = (\theta_{111},\ldots,\theta_{AT2})'$, $\boldsymbol{p} = (p_{111},\ldots,p_{AT2})'$, $\boldsymbol{q} = (q_{111},\ldots,q_{AT2})'$.

We use the griddy Gibbs sampler to draw samples from this joint posterior density. The joint conditional posterior distribution of the missing data have standard multinomial forms. In the joint conditional posterior density, $(\theta_{itj}, p_{itj}, q_{itj})$ are independent over t and j, and they have standard Dirichlet distributions. However, the joint posterior density of $(\mu_0,\mu_1,\mu_2,\mu_3,\tau)$ is not in closed form. Therefore, each of them is obtained using a grid method (e.g., Nandram and Yin (2016)). Also, we need the conditional posterior densities for the incomplete Tables t2-t8, and these are given in Appendix A.

3.2. Inference for finite population proportions

For the i^{th} and area, we assume that a random sample of size n_i is selected from a finite population of size N_i, there is no selection bias, and the n_i selected individuals can be classified into a three-way table of counts.

Our target is the finite population proportion for the j^{th} and row, the k^{th} and column, and ℓ^{th} and length in the i^{th} and area, $P_{ijk\ell}, i = 1,\ldots,A, j = k = \ell = 1,2$. Let N_{it} denote the total number responding for i^{th} and area in the t^{th} and table. Then using standard notation in survey sampling, we can write our target as

$$
\begin{aligned}
&P_{i111} = f_i\bar{z}_{i1} + (1-f_i)\bar{Z}_{i1},\\
&P_{i211} = f_i\bar{z}_{i2} + (1-f_i)\bar{Z}_{i2},\\
&P_{i112} = f_i(\bar{x}_{i1}-\bar{z}_{i1}) + (1-f_i)(\bar{X}_{i1}-\bar{Z}_{i1}),\\
&P_{i212} = f_i(\bar{x}_{i2}-\bar{z}_{i2}) + (1-f_i)(\bar{X}_{i2}-\bar{Z}_{i2}),
\end{aligned}
$$

$$P_{i121} = f_i(\bar{y}_{i1} - \bar{z}_{i1}) + (1 - f_i)(\bar{Y}_{i1} - \bar{Z}_{i1}),$$
$$P_{i221} = f_i(\bar{y}_{i2} - \bar{z}_{i2}) + (1 - f_i)(\bar{Y}_{i2} - \bar{Z}_{i2}),$$
$$P_{i122} = f_i(\bar{w}_{i1} - \bar{x}_{i1} - \bar{y}_{i1} + \bar{z}_{i1}) + (1 - f_i)(\bar{W}_{i1} - \bar{X}_{i1} - \bar{Y}_{i1} + \bar{Z}_{i1}),$$
$$P_{i222} = f_i(\bar{n}_i - \bar{w}_{i1} - \bar{x}_{i2} - \bar{y}_{i2} + \bar{z}_{i2}) + (1 - f_i)(\bar{N}_i - \bar{W}_{i1} - \bar{X}_{i2} - \bar{Y}_{i2} + \bar{Z}_{i2}),$$

where $\bar{z}_{ij}$, $\bar{x}_{ij} - \bar{z}_{ij}$, $\bar{y}_{ij} - \bar{z}_{ij}$, $\bar{w}_{i1} - \bar{x}_{ij} - \bar{y}_{ij} + \bar{z}_{ij}$, and $\bar{n}_i - \bar{w}_{i1} - \bar{x}_{ij} - \bar{y}_{ij} + \bar{z}_{ij}$, $i = 1, \ldots, A$, $j = 1, 2$, are the sample proportions, $\bar{Z}_{ij}$, $\bar{X}_{ij} - \bar{Z}_{ij}$, $\bar{Y}_{ij} - \bar{Z}_{ij}$, $\bar{W}_{i1} - \bar{X}_{ij} - \bar{Y}_{ij} + \bar{Z}_{ij}$, and $\bar{N}_i - \bar{W}_{i1} - \bar{X}_{ij} - \bar{Y}_{ij} + \bar{Z}_{ij}$, $i = 1, \ldots, A$, $j = 1, 2$, are the nonsample proportions, and $f_i = n_i/N_i$, $i = 1, \ldots, A$, are the sampling fractions. Both the sample proportions and the nonsample proportions are unobserved. Thus, given the sampled data, both of them are random variables. While the sample proportion is obtained directly from the model fitting, the nonsample proportion has to be predicted.

Now we show how to predict the nonsampled proportions. Let $\tilde{N}_i = N_i - n_i$ denote the number of nonsample individuals for i^{th} and area, $\tilde{N}_{it} = N_{it} - n_{it}$, $\tilde{\boldsymbol{N}}_i = (\tilde{N}_{i1}, \ldots, \tilde{N}_{i8})'$ and $\boldsymbol{\pi}_i = (\pi_{i1}, \ldots, \pi_{i8})'$. Then under the nonignorable nonresponse model, for $i = 1, \ldots, A$, $t = 2, \ldots, T$, $j = 1, 2$,

$$\tilde{\boldsymbol{N}}_i | \boldsymbol{\pi}_i \sim \text{Multinomial}(N_i - n_i, \boldsymbol{\pi}_i),$$
$$\tilde{N}_{it1} | \tilde{N}_{it}, r_{it} \overset{\text{ind.}}{\sim} \text{Binomial}(\tilde{N}_{it}, r_{it}), \quad \tilde{N}_{it2} = \tilde{N}_{it} - \tilde{N}_{it1},$$
$$Z_{itj}, X_{itj} - Z_{itj}, Y_{itj} - Z_{itj}, W_{itj} - X_{itj} - Y_{itj} + Z_{itj} | \tilde{N}_{itj}, \theta_{itj}, p_{itj}, q_{itj}$$
$$\overset{\text{ind.}}{\sim} \text{Multinomial}\{\tilde{N}_{itj}, (\theta_{itj}, p_{itj} - \theta_{itj}, q_{itj} - \theta_{itj}, 1 - p_{itj} - q_{itj} + \theta_{itj})\}.$$

4. An Illustrative Example and a Simulation Study

We performed two empirical studies to assess the difference between the ignorable nonresponse model and the nonignorable nonresponse models. In Section 4.1 we discuss an illustrative example and in Section 4.2 we describe a simulation study.

4.1. Illustrative example

We have data from the NHANES III (National Center for Health Statistics (1992)) with ten selected states (also called areas). We use three categorical variables; family income divided by family size (INS), number of gardening activities in a week (NG), and bone mineral density (BMD). Because INS indicates poverty, NG shows weekly physical activity volume, and BMD is used to diagnose osteoporosis, a disease of elderly females, we consider the relationship among these

Table 2. Counts and sampling sizes for NHANES III by area.

INS	NG	BMD	Area 1	2	3	4	5	6	7	8	9	10
0	0	0	15	3	1	0	4	1	4	2	2	4
		1	7	0	0	0	0	0	1	0	1	3
		2	4	0	0	0	0	0	0	0	0	6
	1	0	7	1	2	0	0	1	3	1	1	1
		1	6	1	0	0	0	1	1	1	0	1
		2	4	3	0	1	0	0	1	0	1	2
	2	0	42	11	11	0	8	1	29	5	10	33
		1	30	9	7	3	2	2	19	3	4	22
		2	15	6	3	1	1	0	14	1	2	22
1	0	0	49	15	5	3	12	1	12	11	6	12
		1	34	10	0	4	6	1	18	3	8	5
		2	5	1	1	0	5	3	2	0	4	4
	1	0	25	0	7	2	3	4	9	4	7	4
		1	26	6	5	3	8	5	7	4	5	3
		2	9	1	1	1	0	0	3	2	2	1
	2	0	70	24	14	4	10	13	47	17	18	36
		1	101	30	12	9	9	8	61	14	27	38
		2	51	11	4	3	4	8	37	7	20	12
2	0	0	4	0	0	0	1	0	4	1	1	4
		1	3	6	1	2	1	0	2	0	0	1
		2	3	1	0	0	1	1	1	0	2	4
	1	0	2	2	2	0	1	0	1	0	1	1
		1	7	0	1	0	1	2	3	1	0	1
		2	0	0	0	0	1	0	0	0	0	3
	2	0	14	8	4	1	1	1	22	4	2	3
		1	17	6	3	7	1	5	16	2	3	8
		2	9	8	2	4	0	1	9	0	4	8
Sampling size (n)			559	163	86	48	80	59	326	83	131	242

NOTE: INS is 0 if family income divided by family size is less than 4.75, 1 if it is greater than 4.75 and 2 if it is missing. NG is 0 if the number of gardening in a week is less than 4, 1 if it is greater than 4 and 2 if it is missing. Set BMD 0 if BMD is actually greater than 0.64 mg/cm^2 (normal), 1 if it is less than 0.64 mg/cm^2 (bone disease), and 2 if missing. The cutpoints for BMD are set by the WHO, and those for INS and NG are suggested by experts.

variables. In Table 2, we present the contingency tables of counts by state. We also present the sample size (n_i) of the i^{th} and area and, for prediction of the finite population proportions, we take the population sizes $N_i = 20 \times n_i, i = 1, \ldots, 10$. The cutpoints for BMD are set by the World Health Organization (WHO), and those for INS and NG are suggested by experts.

We used the griddy Gibbs sampler to fit the four models to the NHANES III data. For all models, we monitored the convergence of the Gibbs samplers using trace plots, autocorrelation plots and Geweke test of stationarity. For our data, we used 11,000 iterates with the first 1,000 iterates used as a 'burn-in' period (based on the trace plots). We found negligible autocorrelations among the iterates, and so we did not need to have thinning to provide a random sample. In our Gibbs samplers, the p-values of the Geweke test were all greater than 0.10.

First, we fit the ignorable nonresponse model, and we recorded summaries for r_i, θ_i, p_i, q_i and the finite population proportions. When we fit the ignorable nonresponse model, we obtained estimates of $\mu_{00}, \mu_{10}, \mu_{20}, \mu_{30}$ and α_0, β_0. The posterior means and 95% credible intervals (CI) were as follows: For μ_{00}, $\mu_{00}^{(0)} = 0.305$ and $CI = (0.091, 0.656)$; for μ_{10}, $\mu_{10}^{(0)} = 0.298$ and $CI = (0.062, 0.544)$; for μ_{20}, $\mu_{20}^{(0)} = 0.546$ and $CI = (0.290, 0.766)$; for μ_{30}, $\mu_{30}^{(0)} = 0.508$ and $CI = (0.274, 0.756)$; for τ_0, $\tau_0^{(0)} = 18.255$ and $CI = (1.601, 73.032)$. We got values of α_0 and β_0 using maximum likelihood estimates, $\alpha_0^{(0)} = 5.009$ and $\beta_0^{(0)} = 0.125$ for Nig1. For Nig3, we fit the ignorable nonresponse model and recorded summaries for $r_i, \theta_{ij}, p_{ij}, q_{ij}$. The posterior means and 95% credible intervals were as follows: For μ_{00}, $\mu_{00}^{(0)} = 0.312$ and $CI = (0.094, 0.657)$; for μ_{10}, $\mu_{10}^{(0)} = 0.305$ and $CI = (0.078, 0.541)$; for μ_{20}, $\mu_{20}^{(0)} = 0.551$ and $CI = (0.274, 0.778)$; for μ_{30}, $\mu_{30}^{(0)} = 0.518$ and $CI = (0.260, 0.757)$; for τ_0, $\tau_0^{(0)} = 17.974$ and $CI = (1.623, 76.224)$. We got values of α_0 and β_0 using maximum likelihood estimates, $\alpha_0^{(0)} = 6.469$ and $\beta_0^{(0)} = 0.145$ for Nig3.

We compared the ignorable nonresponse model (Ig) and the four nonignorable nonresponse models (Nig1, Nig2, Nig3, Nig4). We studied the finite population proportion corresponding to each cell of the $2 \times 2 \times 2$ categorical table. To investigate the pooling features of the four models, we also fit them to each area separately. We have studied all eight cells of the three-way table by area but, because of space restrictions, we have presented numerical summaries – posterior mean (PM), posterior standard deviation (PSD), and 95% credible interval (CI) – in Table 3 for the cell $(1, 1, 1)$ by state (i.e., P_{111} dropping subscript i).

The PMs are very similar for all the models, and they differ significantly over areas. There are no surprises here. A similar pattern holds for the direct estimates. But it is surprising that the pooled estimates can differ substantially from the direct estimates. Sometimes they are much larger and sometimes much smaller. This phenomenon is likely due to the sparseness of the contingency tables with many of the cell counts too small. The PMs over the pooled models are mostly similar.

Table 3. Posterior inference of the finite population proportion for cell $(1, 1, 1)$ by area.

Area	Model	Direct PM	Direct PSD	Direct CI	Pooled PM	Pooled PSD	Pooled CI
1	Ig	0.085	0.010	(0.067, 0.104)	0.084	0.009	(0.067, 0.102)
	Nig1	0.083	0.014	(0.055, 0.110)	0.082	0.011	(0.061, 0.104)
	Nig2	0.086	0.013	(0.061, 0.114)	0.082	0.012	(0.059, 0.106)
	Nig3	0.091	0.015	(0.062, 0.122)	0.094	0.013	(0.069, 0.121)
	Nig4	0.094	0.016	(0.064, 0.125)	0.093	0.015	(0.066, 0.122)
2	Ig	0.100	0.020	(0.064, 0.142)	0.092	0.017	(0.062, 0.127)
	Nig1	0.085	0.026	(0.034, 0.138)	0.083	0.014	(0.057, 0.112)
	Nig2	0.094	0.027	(0.044, 0.149)	0.083	0.015	(0.055, 0.114)
	Nig3	0.093	0.028	(0.042, 0.151)	0.087	0.016	(0.059, 0.120)
	Nig4	0.097	0.032	(0.038, 0.161)	0.087	0.017	(0.057, 0.122)
3	Ig	0.076	0.026	(0.033, 0.134)	0.088	0.021	(0.050, 0.134)
	Nig1	0.085	0.032	(0.031, 0.155)	0.098	0.019	(0.064, 0.140)
	Nig2	0.102	0.034	(0.044, 0.174)	0.099	0.021	(0.061, 0.144)
	Nig3	0.079	0.032	(0.026, 0.149)	0.103	0.021	(0.065, 0.149)
	Nig4	0.084	0.038	(0.027, 0.174)	0.103	0.022	(0.063, 0.152)
4	Ig	0.023	0.014	(0.004, 0.058)	0.048	0.017	(0.019, 0.086)
	Nig1	0.028	0.016	(0.005, 0.068)	0.067	0.014	(0.041, 0.096)
	Nig2	0.043	0.022	(0.010, 0.095)	0.065	0.015	(0.038, 0.096)
	Nig3	0.028	0.017	(0.004, 0.068)	0.072	0.017	(0.042, 0.108)
	Nig4	0.030	0.020	(0.004, 0.079)	0.072	0.016	(0.044, 0.107)
5	Ig	0.106	0.027	(0.059, 0.164)	0.094	0.021	(0.057, 0.140)
	Nig1	0.103	0.027	(0.057, 0.161)	0.088	0.017	(0.058, 0.124)
	Nig2	0.101	0.029	(0.052, 0.164)	0.088	0.018	(0.056, 0.126)
	Nig3	0.116	0.032	(0.062, 0.186)	0.093	0.019	(0.060, 0.134)
	Nig4	0.122	0.034	(0.064, 0.198)	0.092	0.019	(0.058, 0.134)
6	Ig	0.032	0.017	(0.007, 0.072)	0.052	0.017	(0.022, 0.090)
	Nig1	0.038	0.018	(0.010, 0.080)	0.069	0.014	(0.042, 0.098)
	Nig2	0.044	0.021	(0.014, 0.092)	0.068	0.015	(0.040, 0.101)
	Nig3	0.037	0.018	(0.010, 0.080)	0.071	0.016	(0.042, 0.106)
	Nig4	0.040	0.021	(0.010, 0.090)	0.071	0.016	(0.043, 0.105)
7	Ig	0.081	0.014	(0.054, 0.111)	0.082	0.013	(0.058, 0.110)
	Nig1	0.081	0.019	(0.046, 0.120)	0.085	0.014	(0.059, 0.113)
	Nig2	0.090	0.019	(0.055, 0.130)	0.086	0.015	(0.057, 0.115)
	Nig3	0.088	0.020	(0.053, 0.130)	0.096	0.016	(0.065, 0.129)
	Nig4	0.092	0.023	(0.050, 0.141)	0.096	0.017	(0.064, 0.131)
8	Ig	0.076	0.023	(0.039, 0.126)	0.078	0.019	(0.045, 0.118)
	Nig1	0.077	0.023	(0.037, 0.126)	0.080	0.016	(0.052, 0.114)
	Nig2	0.078	0.024	(0.040, 0.134)	0.080	0.016	(0.051, 0.116)
	Nig3	0.078	0.024	(0.037, 0.130)	0.079	0.017	(0.049, 0.116)
	Nig4	0.080	0.027	(0.036, 0.140)	0.077	0.017	(0.048, 0.115)

Table 3. Continued.

Area	Model	Direct			Pooled		
		PM	PSD	CI	PM	PSD	CI
9	Ig	0.048	0.014	(0.024, 0.080)	0.057	0.014	(0.033, 0.087)
	Nig1	0.052	0.016	(0.025, 0.089)	0.070	0.013	(0.046, 0.098)
	Nig2	0.059	0.017	(0.031, 0.097)	0.069	0.014	(0.043, 0.097)
	Nig3	0.060	0.020	(0.027, 0.106)	0.081	0.016	(0.052, 0.114)
	Nig4	0.064	0.022	(0.028, 0.115)	0.080	0.017	(0.051, 0.116)
10	Ig	0.185	0.026	(0.135, 0.238)	0.163	0.023	(0.121, 0.209)
	Nig1	0.175	0.033	(0.108, 0.241)	0.124	0.021	(0.085, 0.167)
	Nig2	0.180	0.035	(0.111, 0.247)	0.127	0.022	(0.088, 0.173)
	Nig3	0.187	0.032	(0.126, 0.252)	0.132	0.022	(0.092, 0.177)
	Nig4	0.192	0.036	(0.121, 0.261)	0.133	0.024	(0.089, 0.184)

NOTE : PM is the posterior mean, PSD is the posterior standard deviation and CI is the 95% credible interval.

As expected, the PSDs under the pooled models are smaller than those under the individual models, sometimes substantially. For example, in Table 3 for State 2 the PSDs under the pooled (individual) models are 0.017 (0.020), 0.014 (0.026), 0.015 (0.027), 0.017 (0.032). This is a well established result in small area estimation. The PSDs over the pooled models are mostly similar. We have seen states (4, 5, 6, 8) where the PSDs are slightly larger under the ignorable nonresponse model. There are several reasons for this, but we do not discuss them here.

We also drew plots for PM (top panel) and PSD (bottom panel) of the finite population proportions from direct models versus pooled models in Figure 2. We have seen the same patterns in all cells of the three-way contingency tables (seven plots are omitted to save space). We have compared the PMs for the pooled models over states by cell using plots (not shown), and we observed that the four models are generally similar. For each cell we have noticed one or two states in which the estimates under Ig are substantially smaller or higher than under Nig1, Nig2, Nig3, and Nig4. Some examples are P_{121}, P_{211}, P_{221}, and P_{222} for State 2, State 5, State 3, and State 6, respectively.

To assess the overall fit of the models, we performed three goodness-of-fit procedures, the deviance information criterion (DIC) together with the complexity (PD) or effective number of parameters, the Bayesian posterior predictive p-value (BPP), and the log pseudo marginal likelihood (LPML), a summary of the conditional predictive ordinate (CPO) values. For any of the nonignorable nonresponse models (e.g., Nig4), letting $\boldsymbol{d} = (\boldsymbol{d}_{obs}, \boldsymbol{d}_{mis})$,

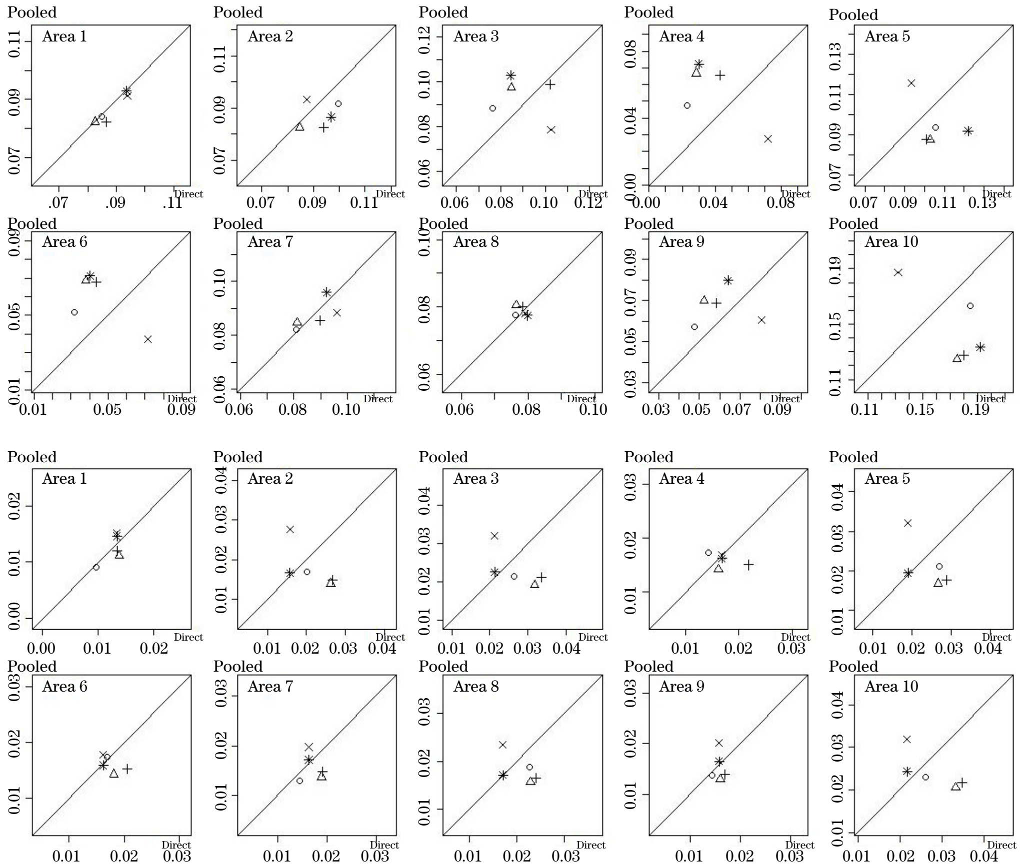

Figure 2. Plots of PMs (top panel) and PSDs (botton panel) of the finite population proportion for cell $(1, 1, 1)$ by area.

NOTE : $\circ$ is Ig, $\triangle$ is Nig1, $+$ is Nig2, $\times$ is Nig3 and $*$ is Nig4.

$$p(\boldsymbol{d}|\boldsymbol{r}, \boldsymbol{\theta}, \boldsymbol{p}, \boldsymbol{q}) = \prod_{i=1}^{A} \prod_{t=1}^{T} \prod_{j=1}^{2} p(w_{it1}, x_{it1}, y_{it1}, z_{it1}, x_{it2}, y_{it2}, z_{it2}|r_{it}, \theta_{itj}, p_{itj}, q_{itj}).$$

Let $r_{it}^{(h)}, \theta_{itj}^{(h)}, p_{itj}^{(h)}, q_{itj}^{(h)}$, $i = 1, \ldots, A$, $t = 1, \ldots, T$, $j = 1, 2$, $h = 1, \ldots, H$, denote the iterates from the griddy Gibbs sampler under the nonignorable nonresponse model and let the posterior means be $\bar{r}_{it}, \bar{\theta}_{itj}, \bar{p}_{itj}, \bar{q}_{itj}$ and let $\boldsymbol{r}^{(h)} = \left(r_{11}^{(h)}, \ldots, r_{AT}^{(h)}\right)'$, $\boldsymbol{\theta}^{(h)} = \left(\theta_{111}^{(h)}, \ldots, \theta_{AT2}^{(h)}\right)'$, $\boldsymbol{p}^{(h)} = \left(p_{111}^{(h)}, \ldots, p_{AT2}^{(h)}\right)'$, $\boldsymbol{q}^{(h)} = \left(q_{111}^{(h)}, \ldots, q_{AT2}^{(h)}\right)'$, $\bar{\boldsymbol{r}} = (\bar{r}_{11}, \ldots, \bar{r}_{AT})'$, $\bar{\boldsymbol{\theta}} = (\bar{\theta}_{111}, \ldots, \bar{\theta}_{AT2})'$, $\bar{\boldsymbol{p}} = (\bar{p}_{111}, \ldots, \bar{p}_{AT2})'$, $\bar{\boldsymbol{q}} = (\bar{q}_{111}, \ldots, \bar{q}_{AT2})'$. Then, for the nonignorable nonresponse model the deviance information criterion is given by

$$DIC = 2\bar{D} - D(\bar{\boldsymbol{r}}, \bar{\boldsymbol{\theta}}, \bar{\boldsymbol{p}}, \bar{\boldsymbol{q}}), \tag{4.1}$$

$\bar{D} = -2\sum_{h=1}^{H} log(p(\boldsymbol{d}|\boldsymbol{r}^{(h)}, \boldsymbol{\theta}^{(h)}, \boldsymbol{p}^{(h)}, \boldsymbol{q}^{(h)}))/H$ and $D(\bar{\boldsymbol{r}}, \bar{\boldsymbol{\theta}}, \bar{\boldsymbol{p}}, \bar{\boldsymbol{q}}) = -2log(p(\boldsymbol{d}|\bar{\boldsymbol{r}}, \bar{\boldsymbol{\theta}}, \bar{\boldsymbol{p}}, \bar{\boldsymbol{q}}))$. Models with smaller DICs are preferred over those with larger DICs. However, since DIC tends to select over-fitted models, Yan and Sedransk (2007) described the Bayesian predictive p-values as a backup.

Let $y_{itjk\ell}$ and $p_{itjk\ell}$ denote, respectively, the cell (j, k, ℓ) count and probability in the t^{th} and table for the i^{th} and area and let $\boldsymbol{y}_{it} = (y_{it111}, \ldots, y_{itrcu})'$ be the multinomial distribution with probabilities $\boldsymbol{p}_{it} = (p_{it111}, \ldots, p_{itrcu})'$. Clearly, $E(y_{itjk\ell}|p_{itjk\ell}) = n_{it}p_{itjk\ell}$ and $Var(y_{itjk\ell}|p_{itjk\ell}) = n_{it}p_{itjk\ell}(1 - p_{itjk\ell})$. For the nonignorable nonresponse model, we choose the discrepancy to be

$$T(\boldsymbol{y}; \boldsymbol{p}) = \sum_{i=1}^{A}\sum_{t=1}^{T}\sum_{jk\ell} \frac{\{y_{itjk\ell} - E(y_{itjk\ell}|p_{itjk\ell})\}^2}{Var(y_{itjk\ell}|p_{itjk\ell})},$$

where $\boldsymbol{y} = (y_{11111}, \ldots, y_{ATrcu})'$ and $\boldsymbol{p} = (p_{11111}, \ldots, p_{ATrcu})'$. Then, we can obtain the respective Bayesian predictive p-values corresponding to the models, $p\{T(\boldsymbol{y}^{(rep)}; \boldsymbol{p}) \geq T(\boldsymbol{y}^{(obs)}; \boldsymbol{p})\}$. Here, these probabilities are calculated over their corresponding iterates $\boldsymbol{p}^{(h)}$, $h = 1, \ldots, H$. A value of this probability, not close to 0 or 1, is indicative of a good fit of the model.

In addition, we can give another measure for evaluating the goodness-of-fit, LPML. The likelihood of each area is given by

$$\begin{aligned} &p(\boldsymbol{w}_i, \boldsymbol{z}_i, \boldsymbol{x}_i, \boldsymbol{y}_i|\boldsymbol{r}_i, \boldsymbol{\theta}_i, \boldsymbol{p}_i, \boldsymbol{q}_i) \\ &= \prod_{t=1}^{T} p(w_{it1}, z_{it1}, x_{it1}, y_{it1}, z_{it2}, x_{it2}, y_{it2}|r_{it}, \theta_{it1}, p_{it1}, q_{it1}, \theta_{it2}, p_{it2}, q_{it2}), \end{aligned}$$

where $\boldsymbol{w}_i = (w_{i11}, \ldots, w_{iT2})'$, $\boldsymbol{z}_i = (z_{i11}, \ldots, z_{iT2})'$, $\boldsymbol{x}_i = (x_{i11}, \ldots, x_{iT2})'$, $\boldsymbol{y}_i = (y_{i11}, \ldots, y_{iT2})'$, $\boldsymbol{r}_i = (r_{i1}, \ldots, r_{iT})'$, $\boldsymbol{\theta}_i = (\theta_{i11}, \ldots, \theta_{iT2})'$, $\boldsymbol{p}_i = (p_{i11}, \ldots, p_{iT2})'$, and $\boldsymbol{q}_i = (q_{i11}, \ldots, q_{iT2})'$. The leave-one-out cross-validation predictive density under the nonignorable nonresponse model is given by

$$\begin{aligned} &p(\boldsymbol{w}_i, \boldsymbol{z}_i, \boldsymbol{x}_i, \boldsymbol{y}_i|\boldsymbol{w}_{(i)}, \boldsymbol{z}_{(i)}, \boldsymbol{x}_{(i)}, \boldsymbol{y}_{(i)}) \\ &= \int p(\boldsymbol{w}_i, \boldsymbol{z}_i, \boldsymbol{x}_i, \boldsymbol{y}_i|\Omega)\pi(\Omega|\boldsymbol{w}_{(i)}, \boldsymbol{z}_{(i)}, \boldsymbol{x}_{(i)}, \boldsymbol{y}_{(i)})d\Omega, \end{aligned}$$

where $(\boldsymbol{w}_{(i)}, \boldsymbol{z}_{(i)}, \boldsymbol{x}_{(i)}, \boldsymbol{y}_{(i)})$ denotes $(\boldsymbol{w}_i, \boldsymbol{z}_i, \boldsymbol{x}_i, \boldsymbol{y}_i)$ after omitting the data from the i^{th} and area and $\Omega = (\boldsymbol{r}, \boldsymbol{\theta}, \boldsymbol{p}, \boldsymbol{q})$. Then, the CPO can be estimated by

$$\widehat{CPO}_i = \left\{\frac{1}{H}\sum_{h=1}^{H} \frac{1}{p\left(\boldsymbol{w}_i, \boldsymbol{z}_i, \boldsymbol{x}_i, \boldsymbol{y}_i\middle|\boldsymbol{r}_i^{(h)}, \boldsymbol{\theta}_i^{(h)}, \boldsymbol{p}_i^{(h)}, \boldsymbol{q}_i^{(h)}\right)}\right\}^{-1},$$

where $\left(\boldsymbol{r}_i^{(h)}, \boldsymbol{\theta}_i^{(h)}, \boldsymbol{p}_i^{(h)}, \boldsymbol{q}_i^{(h)}\right)$ are h^{th} and samples from the griddy Gibbs samper.

Table 4. Model Diagnostic statistics.

Model	DIC	BPP	LPML
Ig	1,489.1	0.2582	−724.5389
Nig1	1,472.6	0.5356	−709.5037
Nig2	1,466.4	0.5385	−701.1168
Nig3	1,471.4	0.4650	−707.4640
Nig4	1,470.5	0.4821	−708.0226

NOTE: Nig2 is selected.

Then LPML is

$$LPML = \sum_{i=1}^{A} \log(\widehat{CPO_i}).$$

Here larger values of LPML indicate better fitting models.

The results of diagnostic statistics are shown in Table 4. The BPPs of the models are not close to 0 or 1 so that, based on the BPPs, there are virtually no differences among the models. DICs of the nonignorable nonresponse models are lower than those of the ignorable nonresponse model and this indicates that the nonognorable nonresponse models are better. Also LPMLs of the nonignorable nonresponse models are larger than the one under the ignorable nonresponse model, again indicating the nonignorable nonresponse models are better than the ignorable nonresponse model. It is interesting that Nig2 appears to be the best model overall. Because Nig1 has a data-based prior, we prefer not to recommend it. While Nig2 is selected by the DIC and LPML, we prefer Nig4 because it is the most general model; the differences in the DICs or LPML are small at any rate.

4.2. Simulation study

We performed a simulation study to further assess the performance between the ignorable nonresponse model (Ig) and the general nonignorable nonresponse model (Nig4). We kept $r = c = \ell = 2$ and $A = 10$ and the sample size n_i as in the original data. After fitting the ignorable nonresponse model to NHANES III data, we obtained the posterior means $\hat{\mu}_0 = 0.2512$, $\hat{\mu}_1 = 0.3269$, $\hat{\mu}_2 = 0.5767$, $\hat{\mu}_3 = 0.5133$ and $\hat{\tau} = 39.0423$ for the hyper parameters μ_0, μ_1, μ_2, μ_3, and τ. Our strategy was to generate data from Ig and fit both Ig and Nig4 and to generate data from Nig4 and fit both Ig and Nig4. Apart from the Ig-Nig4 dichotomy, we also studied the effects of increasing the number of areas (states) (A = 10, 25, 50, 100). Thus, we have an experiment with two factors (model at two levels and number of areas at 4 levels). To get more areas, we replicated the data for the

10 areas (we set only marginal sample sizes).

Thus, we generated cell proportions from (a) the ignorable nonresponse model,

$$r_i|\hat{\mu}_0, \hat{\tau} \overset{\text{i.i.d.}}{\sim} \text{Beta}\{\hat{\mu}_0\hat{\tau}, (1-\hat{\mu}_0)\hat{\tau}\},$$
$$(\theta_i, p_i - \theta_i, q_i - \theta_i, 1 - p_i - q_i + \theta_i)|\hat{\mu}_1, \hat{\mu}_2, \hat{\mu}_3, \hat{\tau} \overset{\text{i.i.d.}}{\sim}$$
$$\text{Dirichlet}\{\hat{\mu}_1\hat{\tau}, (\hat{\mu}_2 - \hat{\mu}_1)\hat{\tau}, (\hat{\mu}_3 - \hat{\mu}_1)\hat{\tau}, (1 - \hat{\mu}_2 - \hat{\mu}_3 + \hat{\mu}_1)\hat{\tau}\}$$

and (b) the nonignorable nonresponse model,

$$r_{it}|\hat{\mu}_0, \hat{\tau} \overset{\text{i.i.d.}}{\sim} \text{Beta}\{\hat{\mu}_0\hat{\tau}, (1-\hat{\mu}_0)\hat{\tau}\},$$
$$(\theta_{itj}, p_{itj} - \theta_{itj}, q_{itj} - \theta_{itj}, 1 - p_{itj} - q_{itj} + \theta_{itj})|\hat{\mu}_1, \hat{\mu}_2, \hat{\mu}_3, \hat{\tau} \overset{\text{i.i.d.}}{\sim}$$
$$\text{Dirichlet}\{\hat{\mu}_1\hat{\tau}, (\hat{\mu}_2 - \hat{\mu}_1)\hat{\tau}, (\hat{\mu}_3 - \hat{\mu}_1)\hat{\tau}, (1 - \hat{\mu}_2 - \hat{\mu}_3 + \hat{\mu}_1)\hat{\tau}\}.$$

With these values, we generated the cell counts for the ignorable nonresponse model from

$$z_{itj}, x_{itj} - z_{itj}, y_{itj} - z_{itj}, w_{itj} - x_{itj} - y_{ijt} + z_{itj}|r_i, \theta_i, p_i, q_i \overset{\text{ind.}}{\sim}$$
$$\text{Multinomial}\{n_i, r_i(\theta_i, p_i - \theta_i, q_i - \theta_i, 1 - p_i - q_i + \theta_i)\}.$$

In a similar manner, letting $r_{it1} = r_{it}$ and $r_{it2} = 1 - r_{it}$, we generated the cell counts for the nonignorable nonresponse model from

$$z_{itj}, x_{itj} - z_{itj}, y_{itj} - z_{itj}, w_{itj} - x_{itj} - y_{ijt} + z_{itj}|r_{it}, \theta_{itj}, p_{itj}, q_{itj} \overset{\text{ind.}}{\sim}$$
$$\text{Multinomial}\{n_i, r_{itj}(\theta_{itj}, p_{itj} - \theta_{itj}, q_{itj} - \theta_{itj}, 1 - p_{itj} - q_{itj} + \theta_{itj})\}.$$

We performed this procedure to get 100 datasets from the ignorable nonresponse model and 100 datasets from the nonignorable nonresponse model. Then we fit each of these datasets using both the ignorable nonresponse model and the nonignorable nonresponse model (Nig4). To compare results between the direct and pooled models, we also fit each of these datasets to the direct (individually) model. We denote the true values of $P_{ijk\ell}$ by $\hat{P}^{(h)}_{ijk\ell}$, obtained from the simulations. Also, we computed the posterior mean of the finite population proportion, $PM^{(h)}_{ijk\ell}$, and the posterior standard deviation, $PSD^{(h)}_{ijk\ell}$, $h = 1, \ldots, 100$, under Ig and Nig4.

We calculated $AB = \sum_{h=1}^{100} AB^{(h)}$ and $PRMSE = (1/100)\sum_{h=1}^{100} PRMSE^{(h)}$, where

$$AB^{(h)} = \sum_{i=1}^{A}\sum_{j=1}^{r}\sum_{k=1}^{c}\sum_{\ell=1}^{u}\left|\hat{P}^{(h)}_{ijk\ell} - PM^{(h)}_{ijk\ell}\right|,$$

Table 5. Average (AVG) and standard error (SE) of the absolute biases (AB) and the posterior root mean squared errors ($PRMSE$) over the 100 runs by area and model.

		AB				$PRMSE$			
		Direct		Pooled		Direct		Pooled	
A	fitted	AVG	SE	AVG	SE	AVG	SE	AVG	SE
10	Ig	2.8141	0.0288	1.7175	0.0236	0.5474	0.0065	0.3509	0.0044
	Nig4	2.5959	0.0253	1.3226	0.0172	0.5514	0.0060	0.2854	0.0034
25	Ig	6.8462	0.0573	4.0057	0.0367	0.8668	0.0101	0.5164	0.0056
	Nig4	6.2029	0.0409	3.0472	0.0303	0.8586	0.0091	0.4112	0.0046
50	Ig	14.3109	0.0962	7.7513	0.0474	1.2599	0.0141	0.7136	0.0073
	Nig4	13.2332	0.0765	5.8750	0.0401	1.2550	0.0133	0.5580	0.0059
100	Ig	28.3993	0.0970	15.3055	0.0497	1.7536	0.0177	0.9991	0.0086
	Nig4	25.9389	0.0867	11.2179	0.0345	1.7478	0.0175	0.7729	0.0063

NOTE: Data were generated from the nonignorable nonresponse model (Nig4), and both Ig and Nig4 were fitted using direct (individually) model and pooled (small area) model.

$$PRMSE^{(h)} = \sqrt{\sum_{i=1}^{A}\sum_{j=1}^{r}\sum_{k=1}^{c}\sum_{\ell=1}^{u}\left\{\left(\hat{P}^{(h)}_{ijk\ell} - PM^{(h)}_{ijk\ell}\right)^2 + PSD^{(h)2}_{ijk\ell}\right\}}$$

represent, respectively, the bias and the posterior root mean squared error corresponding to the h^{th} and dataset. We also calculated the 95% credible interval for each of the 100 simulated runs. We looked at the width ($W^{(h)}_{ijk\ell}$) and the credible incidence ($I^{(h)}_{ijk\ell}$), where $I^{(h)}_{ijk\ell} = 1$ if the 95% credible interval contains the true value $\hat{P}^{(h)}_{ijk\ell}$ and $I^{(h)}_{ijk\ell} = 0$ if the 95% credible interval does not contain the true value $\hat{P}^{(h)}_{ijk\ell}$. For each area and each model, we took the average of these quantities. For example, the estimated probability content of the 95% credible interval is $C = (1/100)\sum_{h=1}^{100} C^{(h)}$, where $C^{(h)} = \sum_{i=1}^{A}\sum_{j=1}^{r}\sum_{k=1}^{c}\sum_{\ell=1}^{u} I^{(h)}_{ijk\ell}/Arcu$.

When data are generated from Ig and both Ig and Nig4 are fitted, we have seen very similar results, with marginal gains of Ig over Nig in terms of bias, posterior mean squared error, and credible interval estimation (probability contents and widths). So it is not necessary to present summaries for data generated from Ig. In Tables 5 and 6, we present summaries when data are generated from Nig4.

First, we discuss Table 5. AB is much smaller for Nig4 (pooled model) over the direct model. It is also true that AB is smaller under Nig4 than Ig, nearly 50% smaller. AB increases with A, the number of small areas. The results for PRMSE show similar comparisons.

Second, we discuss interval estimation in Table 6. C is higher for the pooled model than direct model and closer to the nominal value of 95%. It is also true

Table 6. Summaries of the estimated probability contents (C) and the widths (W) of the 95% credible intervals over the 100 runs by area and model.

		C				W			
		Direct		Pooled		Direct		Pooled	
A	fitted	AVG	SE	AVG	SE	AVG	SE	AVG	SE
10	Ig	0.8459	0.0048	0.8985	0.0045	0.1315	0.0002	0.0960	0.0003
	Nig4	0.9375	0.0033	0.9475	0.0030	0.1555	0.0004	0.0829	0.0004
25	Ig	0.8572	0.0042	0.8979	0.0036	0.1332	0.0011	0.0904	0.0004
	Nig4	0.9462	0.0032	0.9484	0.0016	0.1567	0.0009	0.0772	0.0003
50	Ig	0.8509	0.0024	0.9016	0.0017	0.1333	0.0006	0.0886	0.0002
	Nig4	0.9426	0.0022	0.9513	0.0013	0.1587	0.0006	0.0751	0.0003
100	Ig	0.8467	0.0017	0.9099	0.0015	0.1327	0.0003	0.0889	0.0001
	Nig4	0.9400	0.0013	0.9638	0.0012	0.1556	0.0003	0.0746	0.0002

NOTE: Data were generated from the nonignorable nonresponse model (Nig4), and both Ig and Nig4 were fitted using direct (individually) model and pooled (small area) model.

that C under Nig4 is closer to the 95% nominal value than under Ig. There are minor increases in the coverage as A increases. It is interesting that while Nig4 has coverages closer to 95%, it gives shorter credible intervals. The pooled Nig4 has a much shorter interval (roughly 50% shorter) than the direct and substantially shorter than Ig. Again, there are small changes with A. We have also studied the highest posterior density (HPD) intervals (not shown), and we have seen similar results.

Our simulation study provides confidence in our pooled nonignorable nonresponse model. We have shown rather convincingly that the direct estimators perform badly and Nig4 is much better than Ig in terms of absolute bias, the preferred posterior root mean squared error, which incorporates both squared bias and variance, and coverages and widths of the 95% credible intervals.

5. Concluding Remarks

The purpose of this paper has been to develop a methodology to analyze data from incomplete three-way contingency tables, each table corresponding to a small area. We have studied one ignorable nonresponse model and four nonignorable nonresponse models, constructing each of the nonignorable nonresponse models with a reduced set of nonidentifiable parameters; each of the eight incomplete tables has a set of parameters. We allowed these parameters to share a common effect, thereby passing on the nonidentifiable effects to a manageable set of parameters. Two of our models use Bayesian uncertainty analysis, where we set artificial priors on these hyper parameters. This allows a study of uncertainty about the finite population proportions. The most general model (one of the two models that use Bayesian uncertainty analysis) is selected. Thus, we have extended NW (Nandram and Woo (2015)) to accommodate small areas.

We have presented an illustrative example to estimate the finite population proportions corresponding to the cells of the $2 \times 2 \times 2$ table for the ten US states in NHANES III. We have used the griddy Gibbs sampler for model fitting and performed a model assessment using DIC, BPP, and LPML. We have shown that there are differences among the ignorable nonresponse model and the nonignorable nonresponse models. We have argued that our most general nonignorable nonresponse model is to be preferred for these data.

Our simulation shows rather convincingly that the nonignorable nonresponse model is better than the ignorable nonresponse model. To make this assessment, we have used relative bias, posterior root mean squared error, and coverage. Indeed, this is plausible because the nonignorable nonresponse model contains some degree of difference between the responders and nonresponders.

Acknowledgment

We thank both referees and the Co-Editor for constructive comments, which have led to an improved presentation. The present study was supported in part by grant of Korea of Health and Welfare, Republic of Korea (CIMI-15-03-03). This work was also supported by a grant from the Simons Foundation (#353953, Balgobin Nandram).

Appendix: Some Conditional Posterior distributions for Gibbs sampling

Given the data, $\boldsymbol{d} = (\boldsymbol{d}_{obs}, \boldsymbol{d}_{mis})$, $\mu_0, \mu_1, \mu_2, \mu_3, \tau$ and $r_t, \theta_{tj}, p_{tj}, q_{tj}$ are independent with

$r_{it}|w_{it1}, \mu_0, \tau, \boldsymbol{d} \overset{\text{ind.}}{\sim} \text{Beta}\{w_{it1} + \mu_0\tau, n_{it} - w_{it1} + (1 - \mu_0)\tau\}$,

$\theta_{itj}, p_{itj} - \theta_{itj}, q_{itj} - \theta_{itj}, 1 - p_{itj} - q_{itj} + \theta_{itj}|n_{it}, z_{itj}, x_{itj}, y_{itj}, \mu_1, \mu_2, \mu_3, \tau, \boldsymbol{d} \overset{\text{ind.}}{\sim}$
$\text{Dirichlet}\{z_{itj} + \mu_1\tau, x_{itj} - z_{itj} + (\mu_2 - \mu_1)\tau, y_{itj} - z_{it} + (\mu_3 - \mu_1)\tau,$
$w_{itj} - x_{itj} - y_{itj} + z_{it} + (1 - \mu_2 - \mu_3 + \mu_1)\tau\}$,

where $i = 1, \ldots, A$, $t = 1, \ldots, T$, $j = 1, 2$ throughout. The joint conditional posterior density of $\mu_0, \mu_1, \mu_2, \mu_3, \tau$ is

$$
\begin{aligned}
&p(\mu_0,\mu_1,\mu_2,\mu_3,\tau|\boldsymbol{r},\boldsymbol{\theta},\boldsymbol{p},\boldsymbol{q},\boldsymbol{d})\\
&\propto \frac{\Gamma(\tau)^{3AT}}{(1+\tau)^2}\prod_{i=1}^{A}\prod_{t=1}^{T}\frac{r_{it}^{\mu_0\tau-1}}{\Gamma(\mu_0\tau)}\frac{(1-r_{it})^{(1-\mu_0)\tau-1}}{\Gamma((1-\mu_0)\tau)}\\
&\quad\times\prod_{i=1}^{A}\prod_{t=1}^{T}\prod_{j=1}^{2}\left\{\frac{\theta_{itj}^{\mu_1\tau-1}}{\Gamma(\mu_1\tau)}\frac{(p_{itj}-\theta_{itj})^{(\mu_2-\mu_1)\tau-1}}{\Gamma((\mu_2-\mu_1)\tau)}\frac{(q_{itj}-\theta_{itj})^{(\mu_3-\mu_1)\tau-1}}{\Gamma((\mu_3-\mu_1)\tau)}\right.\\
&\quad\left.\times\frac{(1-p_{itj}-q_{itj}+\theta_{itj})^{(1-\mu_2-\mu_3+\mu_1)\tau-1}}{\Gamma((1-\mu_2-\mu_3+\mu_1)\tau)}\right\},
\end{aligned}
$$

where $0<\mu_0<1, 0<B\le\mu_1<\mu_2,\mu_3<1,\tau>0$. Samples from this joint conditional posterior density can be obtained using the grid method.

We need the conditional posterior densities for Tables t2-t8 for i^{th} and area, $i=1,\ldots,A$.

For Table t2, $x_{i21}+x_{i22}=x_{i2}, y_{i21}+y_{i22}=y_{i2}$ and $z_{i21}+z_{i22}=z_{i2}$ are observed. Then,

$z_{i21}|z_{i2},r_{i2}\sim \text{Binomial}(z_{i2},r_{i2})$,

$x_{i21}-z_{i21}|x_{i2},z_{i21},r_{i2}\sim \text{Binomial}(x_{i2}-z_{i2},r_{i2})$,

$y_{i21}-z_{i21}|y_{i2},z_{i2},r_{i2}\sim \text{Binomial}(y_{i2}-z_{i2},r_{i2})$,

$w_{i21}-x_{i21}-y_{i21}+z_{i21}|n_{i2},x_{i2},y_{i2},z_{i2},r_{i2}\sim \text{Binomial}(n_{i2}-x_{i2}-y_{i2}+z_{i2},r_{i2})$.

For Table t3, w_{i31},y_{i31} and y_{i32} are observed. Then,

$$
\begin{aligned}
z_{i31}|y_{i31},\theta_{i31},q_{i31}&\sim \text{Binomial}\left(y_{i31},\frac{\theta_{i31}}{q_{i31}}\right),\\
x_{i31}-z_{i31}|w_{i31},y_{i31},\theta_{i31},p_{i31},q_{i31}&\sim \text{Binomial}\left(w_{i31}-y_{i31},\frac{p_{i31}-\theta_{i31}}{1-q_{i31}}\right),\\
z_{i32}|y_{i32},\theta_{i32},q_{i32}&\sim \text{Binomial}\left(y_{i32},\frac{\theta_{i32}}{q_{i32}}\right),\\
x_{i32}-z_{i32}|w_{i32},y_{i32},\theta_{i32},p_{i32},q_{i32}&\sim \text{Binomial}\left(w_{i32}-y_{i32},\frac{p_{i32}-\theta_{i32}}{1-q_{i32}}\right).
\end{aligned}
$$

For Table t4, w_{i41},x_{i41} and x_{i42} are observed. Then,

$$
\begin{aligned}
z_{i41}|x_{i41},\theta_{i41},p_{i41}&\sim \text{Binomial}\left(x_{i41},\frac{\theta_{i41}}{p_{i41}}\right),\\
y_{i41}-z_{i41}|w_{i41},x_{i41},\theta_{i41},p_{i41},q_{i41}&\sim \text{Binomial}\left(w_{i41}-x_{i41},\frac{q_{i41}-\theta_{i41}}{1-p_{i41}}\right),\\
z_{i42}|x_{i42},\theta_{i42},p_{i42}&\sim \text{Binomial}\left(x_{i42},\frac{\theta_{i42}}{p_{i42}}\right),
\end{aligned}
$$

$$y_{i42} - z_{i42}|w_{i42}, x_{i42}, \theta_{i42}, p_{i42}, q_{i42} \sim \text{Binomial}\left(w_{i42} - x_{i42}, \frac{q_{i42} - \theta_{i42}}{1 - p_{i42}}\right).$$

For Table t5, only w_{i51} is observed. Then,

$$\begin{aligned}
&(z_{i51}, x_{i51} - z_{i51}, y_{i51} - z_{i51}, w_{i51} - x_{i51} - y_{i51} + z_{i51})'|w_{i51}, \theta_{i51}, p_{i51}, q_{i51} \\
&\sim \text{Multinomial}\{w_{i51}, (\theta_{i51}, p_{i51} - \theta_{i51}, q_{i51} - \theta_{i51}, 1 - p_{i51} - q_{i51} + \theta_{i51})'\}, \\
&(z_{i52}, x_{i52} - z_{i52}, y_{i52} - z_{i52}, w_{i52} - x_{i52} - y_{i52} + z_{i52})'|w_{i52}, \theta_{i52}, p_{i52}, q_{i52} \\
&\sim \text{Multinomial}\{w_{i52}, (\theta_{i52}, p_{i52} - \theta_{i52}, q_{i52} - \theta_{i52}, 1 - p_{i52} - q_{i52} + \theta_{i52})'\}.
\end{aligned}$$

For Table t6, $x_{i61} + x_{i62} = x_{i6}$ is observed. Then,

$$\begin{aligned}
&(z_{i61}, x_{i61} - z_{i61}, z_{i62}, x_{i6} - x_{i61} - z_{i62})'|x_{i6}, r_{i6}, \theta_{i61}, p_{i61}, \theta_{i62}, p_{i62} \\
&\sim \text{Multinomial}\left\{x_{i6}, \left(\frac{r_{i6}\theta_{i61}}{p_{i61}}, \frac{r_{i6}(p_{i61}-\theta_{i61})}{p_{i61}}, \frac{(1-r_{i6})\theta_{i62}}{p_{i62}}, \frac{(1-r_{i6})(p_{i62}-\theta_{i62})}{p_{i62}}\right)'\right\}, \\
&(y_{i61} - z_{i61}, w_{i61} - x_{i61} - y_{i61} + z_{i61}, y_{i62} - z_{i62}, w_{i62} - x_{i62} - y_{i62} + z_{i62})' \\
&\sim \text{Multinomial}\left\{n_{i6} - x_{i6}, \left(\frac{r_{i6}(q_{i61} - \theta_{i61})}{1 - p_{i61}}, \frac{r_{i6}(1 - p_{i61} - q_{i61} + \theta_{i61})}{1 - p_{i61}},\right.\right. \\
&\left.\left.\frac{(1 - r_{i6})(q_{i62} - \theta_{i62})}{1 - p_{i62}}, \frac{(1 - r_{i6})(1 - p_{i62} - q_{i62} + \theta_{i62})}{1 - p_{i62}}\right)'\right\}.
\end{aligned}$$

For Table t7, $y_{i71} + y_{i72} = y_{i7}$ is observed. Then,

$$\begin{aligned}
&(z_{i71}, y_{i71} - z_{i71}, z_{i72}, y_{i7} - y_{i71} - z_{i72})'|y_{i7}, r_{i7}, \theta_{i71}, p_{i71}, q_{i71}, \theta_{i72}, p_{i72}, q_{i72} \\
&\sim \text{Multinomial}\left\{y_{i7}, \left(\frac{r_{i7}\theta_{i71}}{q_{i71}}, \frac{r_{i7}(q_{i71}-\theta_{i71})}{q_{i71}}, \frac{(1-r_{i7})\theta_{i72}}{q_{i72}}, \frac{(1-r_{i7})(q_{i72}-\theta_{i72})}{q_{i72}}\right)'\right\}, \\
&(x_{i71} - z_{i71}, w_{i71} - x_{i71} - y_{i71} + z_{i71}, x_{i72} - z_{i72}, w_{i72} - x_{i72} - y_{i72} + z_{i72})' \\
&\sim \text{Multinomial}\left\{n_{i7} - y_{i7}, \left(\frac{r_{i7}(p_{i71} - \theta_{i71})}{1 - q_{i71}}, \frac{r_{i7}(1 - p_{i71} - q_{i71} + \theta_{i71})}{1 - q_{i71}},\right.\right. \\
&\left.\left.\frac{(1 - r_{i7})(p_{i72} - \theta_{i72})}{1 - q_{i72}}, \frac{(1 - r_{i7})(1 - p_{i72} - q_{i72} + \theta_{i72})}{1 - q_{i72}}\right)'\right\}.
\end{aligned}$$

For Table t8, all counts are missing. Then,

$$\begin{aligned}
&w_{i81}|r_{i8} \sim \text{Binomial}(n_{i8}, r_{i8}), \\
&(z_{i81}, x_{i81} - z_{i81}, y_{i81} - z_{i81}, w_{i81} - x_{i81} - y_{i81} + z_{i81})'|w_{i81}, \theta_{i81}, p_{i81}, q_{i81} \\
&\sim \text{Multinomial}\{w_{i81}, (\theta_{i81}, p_{i81} - \theta_{i81}, q_{i81} - \theta_{i81}, 1 - p_{i81} - q_{i81} + \theta_{i81})'\}, \\
&(z_{i82}, x_{i82} - z_{i82}, y_{i82} - z_{i82}, w_{i82} - x_{i82} - y_{i82} + z_{i82})'|w_{i82}, \theta_{i82}, p_{i82}, q_{i82} \\
&\sim \text{Multinomial}\{w_{i82}, (\theta_{i82}, p_{i82} - \theta_{i82}, q_{i82} - \theta_{i82}, 1 - p_{i82} - q_{i82} + \theta_{i82})'\}.
\end{aligned}$$

References

Barker, S. G., Rosenberger, W. F. and DerSimonian, R. (1992). Closed-form estimates for missing counts in two-way contingency tables. *Stat. Med.* **11**, 643–657.

Chen, T. and Fienberg, S. E. (1974). Two-dimensional contingency tables with both completely and partially cross-classified data. *Biometrics* **30**, 629–642.

Draper, D. (1995). Assessment and propagation of model uncertainty. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **57**, 45–97.

Forster, J. J. and Smith, P. W. F. (1998). Model-based inference for categorical survey data subject to non-ignorable nonresponse. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **60**, 57–70.

Greenland, S. (2009). Relaxation penalties and priors for plausible modeling of nonidentified Bias Sources. *Statist. Sci.* **24**, 195–210.

Little, R. J. A. and Rubin, D. B. (2002). *Statistical Analysis with Missing Data.* 2nd Edition. Wiley.

Molenberghs, G., Goetghebeur, E., Lipsitz, S. R. and Kenward, M. G. (1999). Nonrandom missingness in categorical data: Strengths and limitations. *Amer. Statist.* **53**, 110–118.

Molenberghs, G., Kenward, M. G. and Goetghebeur, E. (2001). Sensitivity analysis for incomplete contingency table: The Slovenian plebiscite case. *J. R. Stat. Soc. Ser. C. Appl. Stat.* **50**, 15–29.

Nandram, B. and Choi, J. W. (2002). Hierarchical Bayesian nonresponse models for binary data from small area with uncertainty about ignorability. *J. Amer. Statist. Assoc.* **97**, 381–388.

Nandram, B. and Choi, J. W. (2010). A Bayesian analysis of body mass index data from small domains under nonignorable nonresponse and selection. *J. Amer. Statist. Assoc.* **105**, 120–135.

Nandram, B., Cox, L. and Choi, J. W. (2005). Bayesian analysis of nonignorable missing categorical data: An application to bone mineral density and family income. *Surv. Methodol.* **31**, 213–225.

Nandram, B. and Katzoff, M. (2012). A hierarchical Bayesian nonresponse model for two-way categorical data from small areas with uncertainty about ignorability. *Surv. Methodol.* **38**, 81–93.

Nandram, B., Liu, N., Choi, J. W. and Cox, L. (2005). Bayesian nonresponse models for categorical data from small areas: An application to BMD and age. *Stat. Med.* **24**, 1047–1074.

Nandram, B. and Woo, N. K. (2015). A Bayesian uncertainty analysis for nonignorable nonresponse. In *Current Trends in Bayesian Methodology with Applications*, 415–433. (Edited by S. K. Upadhyay, U. Singh, D. K. Dey and A. Longanathan), Chapman and Hall/CRC, Boca Raton.

Nandram, B. and Yin, J. (2016). A nonparametric Bayesian prediction interval for a finite population mean. *J. Stat. Comput. Simul.* **86**, 3141–3157.

National Center for Health Statistics (1992). Third national health and nutrition examination survey. *Vital and Health Stat.* Series **2**, 113.

Rubin, D. B., Stern, H. S. and Behovar, V. (1995). Handling "Don't Know" survey responses: The case of the Slovenian Plebiscite. *J. Amer. Statist. Assoc.* **72**, 822–828.

Yan, G. and Sedransk, J. (2007). Bayesian diagnostic techniques for detecting hierarchical structure. *Bayesian Anal.* **2**, 735–760.

Research & Development Office, Korea Appraisal Board, Daegu 41068, Korea.
E-mail: namkyo.woo@gmail.com
Department of Mathematical Sciences, Worcester Polytechnic Institute, Worcester, MA 01609, USA.
E-mail: balnan@wpi.edu
Department of Statistics, Kyungpook National University, Daegu 41566, Korea.
E-mail: dalkim@knu.ac.kr

(Received March 2016; accepted March 2017)