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

Article

A hierarchical Bayesian nonresponse model for two-way categorical data from small areas with uncertainty about ignorability

by Balgobin Nandram and Myron Katzoff

June 2012

Statistics Canada Statistique Canada

Canada

Survey Methodology, June 2012
Vol. 38, No. 1, pp. 81-93
Statistics Canada, Catalogue No. 12-001-X

A hierarchical Bayesian nonresponse model for two-way categorical data from small areas with uncertainty about ignorability

Balgobin Nandram and Myron Katzoff [1]

Abstract

We study the problem of nonignorable nonresponse in a two dimensional contingency table which can be constructed for each of several small areas when there is both item and unit nonresponse. In general, the provision for both types of nonresponse with small areas introduces significant additional complexity in the estimation of model parameters. For this paper, we conceptualize the full data array for each area to consist of a table for complete data and three supplemental tables for missing row data, missing column data, and missing row and column data. For nonignorable nonresponse, the total cell probabilities are allowed to vary by area, cell and these three types of "missingness". The underlying cell probabilities (*i.e.*, those which would apply if full classification were always possible) for each area are generated from a common distribution and their similarity across the areas is parametrically quantified. Our approach is an extension of the selection approach for nonignorable nonresponse investigated by Nandram and Choi (2002a, b) for binary data; this extension creates additional complexity because of the multivariate nature of the data coupled with the small area structure. As in that earlier work, the extension is an expansion model centered on an ignorable nonresponse model so that the total cell probability is dependent upon which of the categories is the response. Our investigation employs hierarchical Bayesian models and Markov chain Monte Carlo methods for posterior inference. The models and methods are illustrated with data from the third National Health and Nutrition Examination Survey.

Key Words: Metropolis-Hastings sampler; SIR algorithm; Nonignorable nonresponse model; Expansion model.

1. Introduction

In sample surveys, data are typically summarized in two-way categorical tables. We consider the problem of non-ignorable nonresponse for many $r \times c$ categorical tables, each obtained from a single area. For many of these surveys, there are missing data and this gives rise to partial classification of the sampled individuals. Thus, for each two-way table there are both item nonresponse (one of the two categories is missing) and unit nonresponse (both categories are missing). One may not know how the data are missing and a model that includes some difference between the observed data and missing data (*i.e.*, nonignorable missing data) may be preferred. For a general $r \times c$ categorical table, we address the issue of estimation of the cell probabilities of the two-way tables when there is possibly nonignorable nonresponse but there is really no information about ignorability. In such a situation, we would like to express a degree of uncertainty about ignorability. Nandram and Choi (2002a, b) have described an expansion model appropriate for binary data when there are data from many small areas. We will extend this work to $r \times c$ categorical tables.

Letting x denote the covariates and y the response variable, Little and Rubin (2002) describe three types of missing-data mechanism. These types differ according to whether the probability of response (a) is independent of x and y; (b) depends on x but not on y; or (c) depends on y and possibly x. The missing data are missing completely at random (MCAR) in (a), missing at random (MAR) in (b) and the data are missing not at random (MNAR) in (c). Models for MCAR and MAR missing-data mechanisms are called ignorable if the parameters of the dependent variable and the response variable are distinct (Rubin 1976). Models for MNAR missing-data mechanisms are called non-ignorable. The general difficulty with nonignorable non-response model is that the parameters are not identifiable [*e.g.*, see Nandram and Choi (2004, 2005, 2008, 2010) and Nandram, Han and Choi (2002)].

For a $r \times c$ categorical table, let $I_{ijkl} = 1$ if the l^{th} individual within the i^{th} area falls in the j^{th} row and k^{th} column and 0 otherwise. Also, let $J_{il} = 1$ if the l^{th} individual within the i^{th} area has complete information and 0 otherwise. Finally, let $P(J_{il} = 1 \mid I_{ijkl} = 1, I_{ij'k'l} = 0, j' \neq j, k' \neq k) = \pi_{ijk}$. For unit nonresponse, if $\pi_{ijk} = \pi_i$, the model is ignorable; for item nonresponse, if the columns are missing, row is observed and $\pi_{ijk} = \pi_{ij}$ (or $\pi_{ijk} = \pi_i$), the model is ignorable; and if the rows are missing but columns are observed and $\pi_{ijk} = \pi_{ik}$ (or $\pi_{ijk} = \pi_i$), the model is ignorable. All other models are nonignorable; see Rubin (1976) for further explanation.

Nandram and Choi (2002a, b) use an expansion model to study nonignorable nonresponse binary data. The expansion

1. Balgobin Nandram, Department of Mathematical Sciences, Worcester Polytechnic Institue, 100 Institue Road, Worcester, MA 01609. E-mail: balnan@wpi.edu; Myron Katzoff, Office of research and Methodology, National Center for Health Statisitcs, CDC, 3311 Toledo Road, Hyattsville, MD 20782. E-mail: mjk5@cdc.gov.

model, a nonignorable nonresponse model, degenerates into an ignorable nonresponse model (in the spirit of Draper 1995) when a centering parameter is set to unity. This permits an expression of uncertainty about ignorability; see also Forster and Smith (1998).

We discuss the model of Nandram and Choi (2002a, b) for binary data from small areas. So that J_{il} denote the response indicators and I_{il} denote the binary response. Specifically, introducing the centering parameters γ_i for area i to incorporate uncertainty about ignorability, the model of Nandram and Choi (2002a, b) is

$$\begin{aligned} I_{il} \mid p_i &\overset{\text{iid}}{\sim} \text{Bernoulli}(p_i), \\ J_{il} \mid \{\pi_i, J_{il} = 0\} &\overset{\text{iid}}{\sim} \text{Bernoulli}(\pi_i), l = 1, \ldots, n_i, i = 1, \ldots, l, \\ J_{il} \mid \{\pi_i, \gamma_i, y_{il} = 1\} &\overset{\text{iid}}{\sim} \text{Bernoulli}(\gamma_i \pi_i),\ 0 < \gamma_i \pi_i < 1. \end{aligned}$$

When $\gamma_i = 1,$ the nonignorable nonresponse model degenerates to an ignorable nonresponse model. Here γ_i is the ratio of the odds of success among respondents to the odds of success among all individuals for the i^{th} area. The parameter γ_i describes the extent of nonignorability of the response mechanism for area i, and it is through the γ_i that uncertainty about ignorability is incorporated. Nandram and Choi (2002a, b) define $\delta_i = \pi_i\{\gamma_i p_i + (1 - p_i)\}$ to be the probability that an individual responds in area i in the entire population, and with a belief that all the areas are similar they take $(p_i, \delta_i, \gamma_i)$ to have a common distribution. A priori they take beta distributions for p_i and π_i respectively.

Here, the parameters are not identifiable. However, if $\gamma_i = 1,$ then all the parameters are identifiable. That is, identifiability of the parameters depend on the γ_i. Note, that when $\gamma_i = 1,$ we get an ignorable model for a MAR mechanism. As the parameters are identifiable in this model, it is quite sensible to use this model (or similar models) as a baseline model. However, note this model is still not justified because it assumes that missing data are like observed data. Thus, to add flexibility to this ignorable nonresponse model, we use the γ_i.

Let γ_{iuv} be the number of individuals with $I_{il} = u,$ $J_{il} = v\,(u, v = 0, 1)$ in the i^{th} area. Then, under the model,

$$\begin{aligned} (y_{i00}, y_{i01}, y_{i10}, y_{i11}) \mid \pi_i, p_i, \gamma_i &\overset{\text{ind}}{\sim} \text{Multinomial}\{n_i, \\ &(1 - p_i)(1 - \pi_i), (1 - p_i)\ \pi_i, (1 - \gamma_i \pi_i)\ p_i, \gamma_i \pi_i p_i\} \end{aligned}$$

with independence over areas. Here, only y_{i01} and y_{i11} are observed, and therefore all parameters are nonidentifiable if the γ_i are unknown. We obtain the likelihood function in a similar manner for the more complete $r \times c$ categorical table with missing data.

We start with a gamma distribution, and to permit centering on the ignorable nonresponse model, we must take each γ_i to have mean 1. However, we need to use a truncated gamma distribution because $0 < \pi_i < 1$ and $0 < \gamma_i \le 1 / \pi_i$. An interesting idea of Nandram and Choi (2002a, b) is to model the centering as a truncated gamma

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

The model is complete with noninformative prior densities on all hyperparameters. One can use alternative distributions (*e.g.*, a truncated lognormal density) for the $\gamma_i,$ but this is not a key issue and it would not matter much.

One can use an area level model with random effects in which, conditional on the observed data, the nonresponse is dependent upon area-level random effects. This can be formulated using a logit link function, but we have not developed our models in this direction partly because we are not using covariates here; see Nandram and Choi (2010) for the use of covariates and random effects.

The approach in Nandram and Choi (2002a, b) is attractive, but it does not apply immediately to the current $r \times c$ categorical table problem. Specifically, only one centering parameter per area is needed in Nandram and Choi (2002a, b). In our formulation, one now needs rc centering parameters per area; each of these parameters has to have a distribution centered at one to allow degeneration to the ignorable nonresponse model. There are also inequality constraints that must be included in the nonignorable nonresponse model. In addition, one cannot rule out the possibility that these parameters are correlated. The methodology needed to apply the work of Nandram and Choi (2002a, b) to the $r \times c$ categorical table is not straightforward. Noting these difficulties Nandram, Liu, Choi and Cox (2005) (with a single supplemental table) and Nandram, Cox and Choi (2005) (with the three supplemental tables) use a simpler idea, but not quite as elegant as in Nandram and Choi (2002a, b), for centering; see also Nandram and Choi (2005).

Essentially, Nandram, Cox and Choi (2005) and Nandram, Liu, Cox and Choi (2005) assume an ignorable model, obtain samples of the response probabilities and use these sampled response probabilities to fit the response probabilities of a nonignorable nonresponse model while "controlling" its parameters. Of course, a possible alternative occurs when there is information about the degree of nonignorability. However, the problem of incorporating prior information about a systematic departure from ignorability is more complex for our problem, and it would need additional costly field work to obtain such information.

We discuss our philosophy about the nonignorable nonresponse problem, a fundamentally aliased problem. In fact, this problem is extremely difficult and we believe that

there is really no solution to it, but we must try. Without any information one cannot tell how respondents and nonrespondents differ. An ignorable nonresponse model is short because it assumes that respondents and nonrespondents are similar, but the respondents and nonrespondents may differ. Statisticians must not confront imprecision (sampling error) only, but they must be bold enough to study subjectivity (ignorance arising from missing information). Unfortunately, as is well known, nonignorable nonresponse models have nonidentifiable parameters. We discuss how the key nonignorability parameters are identified. We know that if the respondents and nonrespondents are similar, then the γ_i are equal unity, and we get the ignorable nonresponse model with all parameters identified. We can now expand the ignorable nonresponse model into a nonignorable nonresponse model by putting a distribution on these γ_i centered at 1, still maintaining identifiability. One can formulate a nonignorable nonresponse model to add flexibility to the ignorable nonresponse model as we have done in our work; the flexibility is a form of sensitivity analysis, coherent in this case, and indeed it is a Bayesian uncertainty (risk) assessment (*e.g.*, Greenland 2009). This is what we have been doing or trying to do in our work.

In this paper we attempt to solve the difficult problem of Nandram and Choi (2002a, b) in its original form for $r \times c$ tables for many areas. The plan of this paper is as follows. In Section 2, we describe the hierarchical Bayesian model. Specifically we describe the nonignorable nonresponse mechanism and we construct an appropriate prior distribution. In Section 3, we show how to fit the model using the sampling importance resampling (SIR) algorithm to subsample from an approximate posterior density after an innovative collapsing of the complete joint posterior density. In Section 4, we illustrate our methodology with public-use data from thirteen states in the third National Health and Nutrition Examination Survey (NHANES III). Section 5 has concluding remarks.

2. The nonignorable nonresponse model

For the problem of nonresponse in a two-dimensional table, we can have both item and unit nonresponse. Thus, one may consider the full data array to consist of four tables: one for complete data and three supplemental tables - one for missing row information, one for missing column information and a table for which neither row nor column membership has been recorded. Throughout this paper, we index rows by $j = 1, \ldots, r$; columns, by $k = 1, \ldots, c$; and the four tables by $s = 1, 2, 3, 4$. We index areas by $i = 1, 2, \ldots, A$ and individuals within areas by $l = 1, 2, \ldots, n_i$. We next describe the nonignorable nonresponse model (*i.e.*, the expansion model).

2.1 Sampling process

We adapt the terminology and definitions used in Nandram, Cox and Choi (2005) to our situation. For sample individual l in area i, let

$$I_{ijkl} = \begin{cases} 1, & \text{if the outcome category is } (j, k) \\ 0, & \text{otherwise,} \end{cases}$$

and let $\boldsymbol{J}_{il}$ denote one of the 4-tuples $(1, 0, 0, 0)$, $(0, 1, 0, 0)$, $(0, 0, 1, 0)$, $(0, 0, 0, 1)$. We assume that

$$\boldsymbol{I}_{il} \stackrel{\text{def}}{=} \text{vec}(\{I_{ijkl} \mid j = 1, \ldots, r; \; k = 1, \ldots, c\}) \mid \boldsymbol{p}_i \stackrel{\text{iid}}{\sim} \text{Mult}\{1, \boldsymbol{p}_i\} \quad (1)$$

and

$$\boldsymbol{J}_{il} \mid \{I_{ijkl} = 1, \; I_{ij'k'l} = 0 \text{ for all } j' \neq j \text{ and } k' \neq k \mid \boldsymbol{\pi}_{ijk}\} \stackrel{\text{iid}}{\sim} \text{Mult}\{1, \boldsymbol{\pi}_{ijk}\}, \quad (2)$$

where $\boldsymbol{p}_i \stackrel{\text{def}}{=} \text{vec}(\{p_{ijk} \mid j = 1, 2, \ldots, r; k = 1, 2, \ldots, c\})$ is a vector of probabilities for the table of rc categories for the variable of observation which must sum to one and, for cell (j, k) in that two-dimensional table,

$$\boldsymbol{\pi}_{ijk} \stackrel{\text{def}}{=} \text{vec}(\{\pi_{isjk}\} \text{ for } s = 1,2,3,4)$$

is a vector of probabilities which must sum to one.

Next, we define cell counts y_{isjk}, for each table $s = 1, \ldots, 4$ for area i such that, for cell (j, k),

$$(y_{i1\,jk}, y_{i2\,jk}, y_{i3\,jk}, y_{i4\,jk}) = \sum_{l=1}^{n_i} I_{ijkl} \, \boldsymbol{J}_{il},$$

where $y_{i1\,jk}$ are observed and y_{isjk}, for $s = 2, 3, 4$, are latent variables which satisfy the observed constraints $\sum_k y_{i2\,jk} = u_{ij}$, $\sum_j y_{i3\,jk} = v_{ik}$ and $\sum_{j,k} y_{i4\,jk} = w_i$. All inferences will be conditional on the observed quantities, u_{ij}, v_{ik} and w_i. But see Nandram (2009) for the analysis of a single $r \times c$ table under nonresponse when the margins are also random. We will denote the vector of the $y_{i1\,jk}$ by $\boldsymbol{y}_1$, the vector of the y_{isjk}, $s = 2, 3, 4$, by $\boldsymbol{y}_{(1)}$, and the complete vector by $\boldsymbol{y} = (\boldsymbol{y}_1, \boldsymbol{y}_{(1)})'$.

The parameters π_{isjk} are not identifiable. If the distributions of these parameters are known completely, then the nonidentifiability will disappear. Thus, the key issue is how to identify these parameters. We know that if the respondents and nonrespondents are similar (*i.e.*, the four patterns, complete and partially complete tables), then we can take $\pi_{isjk} = \pi_{is}$; this is the ignorable nonresponse model. The π_{is} can be estimated by the proportions of cases falling in the four tables for each area. This is a natural point

to start. To expand the ignorable nonresponse model into a nonignorable model, and still maintain identifiability, first we need a simplification. We take $\pi_{ijks} = \psi_{ijk}\,\pi_{is}$, which gives a nonignorable nonresponse model in which the parameters ψ_{ijk} are not identifiable.

To center the nonignorable model on the ignorable model, we take

$$\pi_{isjk} = \begin{cases} \tilde{\psi}_{ijk}\pi_{is}, & \text{for } s=1, \\ \psi_{ijk}\pi_{is}, & \text{for } s=2,3,4, \end{cases} \tag{3}$$

and require that $\sum_{s=1}^{4}\pi_{is}=1$. A little algebra then yields the relationship

$$\begin{aligned}\tilde{\psi}_{ijk}\pi_{i1} &= \left[1+(1-\psi_{ijk})\left(\frac{1-\pi_{i1}}{\pi_{i1}}\right)\right]\pi_{i1} \\ &= a_{ijk}(\pi_{i1},\psi_{ijk})\,\psi_{ijk}\pi_{i1},\end{aligned} \tag{4}$$

where $a_{ijk}(\pi_{i1},\psi_{ijk}) = \{\psi_{ijk}^{-1}+(\psi_{ijk}^{-1}-1)(\pi_{i1}^{-1}-1)\}$, from which it is clear that $\tilde{\psi}_{ijk}=1$ if, and only if, $\psi_{ijk}=1$. Note that since $0\le\pi_{isjk}\le 1$ and $(1-\pi_{i1})^{-1}\le\min\{\pi_{is}^{-1}; s=2,3,4\}$, it follows that $0<\psi_{ijk}\le(1-\pi_{i1})^{-1}$.

By combining (1) and (2) and noting the definition of π_{isjk} in (3), similar to binary case, we get a multinomial distribution for $\boldsymbol{y}$ conditional on $\boldsymbol{\pi}, \boldsymbol{\psi}, \boldsymbol{p}$, and the likelihood function for the sample can now be seen to be

$$\begin{aligned} &f(\boldsymbol{y}\,|\,\boldsymbol{\pi},\boldsymbol{\psi},\boldsymbol{p}) \\ &= \prod_{i=1}^{A}\binom{n_i}{\boldsymbol{y}'_{i1},\boldsymbol{y}'_{i2},\boldsymbol{y}'_{i3},\boldsymbol{y}'_{i4}}\left\{\left[\prod_{j,k}(\tilde{\psi}_{ijk}\pi_{i1}p_{ijk})^{y_{i1jk}}\right]\right. \\ &\qquad\left.\prod_{s=2}^{4}\prod_{j,k}(\psi_{ijk}\pi_{is}p_{ijk})^{y_{isjk}}\right\} \\ &= \prod_{i=1}^{A}\binom{n_i}{\boldsymbol{y}'_{i1},\boldsymbol{y}'_{i2},\boldsymbol{y}'_{i3},\boldsymbol{y}'_{i4}}\left\{\prod_{s=1}^{4}\prod_{j,k}(\psi_{ijk}\pi_{is}p_{ijk})^{y_{isjk}}\right. \\ &\qquad\left.\prod_{j,k}\left[a_{ijk}(\pi_{i1},\psi_{ijk})\right]^{y_{i1jk}}\right\}, \end{aligned} \tag{5}$$

where

$$\begin{aligned} \boldsymbol{y}_{is}^{rc\times 1} &= \mathrm{vec}(\{y_{isjk}\mid j=1,\ldots,r;k=1,\ldots,c\}), \\ \boldsymbol{y} &= (\boldsymbol{y}'_{11},\boldsymbol{y}'_{12},\boldsymbol{y}'_{13},\boldsymbol{y}'_{14},\boldsymbol{y}'_{21},\ldots,\boldsymbol{y}'_{24},\ldots,\boldsymbol{y}'_{A1},\boldsymbol{y}'_{A2},\boldsymbol{y}'_{A3},\boldsymbol{y}'_{A4})', \\ \boldsymbol{\pi}^{A4\times 1} &= (\pi_{11},\ldots,\pi_{14},\pi_{21},\ldots,\pi_{24},\ldots,\pi_{A1},\ldots,\pi_{A4})', \\ \boldsymbol{\psi}^{Arc\times 1} &= (\psi_{111},\ldots,\psi_{1rc},\psi_{211},\ldots\psi_{2rc},\ldots,\psi_{A11},\ldots,\psi_{Arc}), \end{aligned}$$

and

$$\boldsymbol{p}^{Arc\times 1} = (\boldsymbol{p}_1,\boldsymbol{p}_2,\ldots,\boldsymbol{p}_A)'.$$

Collecting factors which are powers of π_{is}, the likelihood function may also be expressed as

$$\begin{aligned} f(\boldsymbol{y}\,|\,\boldsymbol{\pi},\boldsymbol{\psi},\boldsymbol{p}) &= \prod_{i=1}^{A}\binom{n_i}{\boldsymbol{y}'_{i1},\boldsymbol{y}'_{i2},\boldsymbol{y}'_{i3},\boldsymbol{y}'_{i4}} \\ &\left\{\prod_{s=1}^{4}\pi_{is}^{y_{is..}}\times\prod_{j,k}\{p_{ijk}\,\psi_{ijk}\}^{y_{i.jk}}\left[a_{ijk}(\pi_{i1},\psi_{ijk})\right]^{y_{i1jk}}\right\}, \end{aligned} \tag{6}$$

where $0\le\pi_{is}\le 1, \sum_s\pi_{is}=1$ and $0\le\psi_{ijk}\le(1-\pi_{i1})^{-1}$. Here we note that $y_{is..}$ and y_{i1jk} are observed variables but the $y_{i.jk}$ are latent variables.

2.2 Prior construction

The following assumptions describe the prior distributions for the nonignorable nonresponse model:

1. For the vector of cell probabilities $\boldsymbol{p}_i$, we assume that

$$\boldsymbol{p}_i\,|\,\boldsymbol{\mu}_1,\tau_1 \overset{\text{iid}}{\sim} \text{Dirichlet}(\boldsymbol{\mu}_1\tau_1),$$

where $\mu_1=(\mu_{111},\mu_{112},\ldots,\mu_{11k},\mu_{121},\ldots,\mu_{1rc})'$; $\mu_{1jk}\ge 0$; and $\sum_{j=1}^{r}\sum_{k=1}^{c}\mu_{1jk}=1$. The parameter τ_1 informs us of similarity among the $\boldsymbol{p}_i$: the larger τ_1, the more alike the $\boldsymbol{p}_i$. This is true because large τ_1 means that the variances of the $\boldsymbol{p}_i$ are small, and because they have the same mean, this means that they are more similar with larger τ_1.

Thus, the density for $\boldsymbol{p}$ is

$$\begin{aligned} g_1(\boldsymbol{p}\,|\,\boldsymbol{\mu}_1,\tau_1) &= \prod_{i=1}^{A}g_{1i}(\boldsymbol{p}_i\,|\,\boldsymbol{\mu}_1,\tau_1) \\ &= \prod_{i=1}^{A}\left\{\frac{\prod_{j,k}p_{ijk}^{\mu_{1jk}\tau_1-1}}{\mathrm{D}(\boldsymbol{\mu}_1\tau_1)}\right\}, \end{aligned} \tag{7}$$

where, for a k-tuple $\boldsymbol{c}$ and a scalar t

$$\mathrm{D}(\boldsymbol{c}t) = \frac{\prod_{j=1}^{k}\Gamma(\boldsymbol{c}_j t)}{\Gamma(t)}$$

for $c_j>0$ and $\sum_{j=1}^{k}c_j=1$.

2. Independently of the $\boldsymbol{p}_i$, the $\boldsymbol{\pi}_i=(\pi_{i1},\pi_{i2},\pi_{i3},\pi_{i4})'$ follow the specification

$$\boldsymbol{\pi}_i \overset{\text{iid}}{\sim} \text{Dirichlet}(\boldsymbol{\mu}_2\tau_2),$$

with $\pi_{is}\ge 0$ and $\sum_s\pi_{is}=1$, where $\boldsymbol{\mu}_2=(\mu_{21},\mu_{22},\mu_{23},\mu_{24})'$, $\mu_{2s}\ge 0, \sum_{s=1}^{4}\mu_{2s}=1$ and τ_2 is a measure of similarity among the $\boldsymbol{\pi}_i$. Thus, the density for $\boldsymbol{\pi}_i$ is

$$g_{2i}(\boldsymbol{\pi}_i\,|\,\boldsymbol{\mu}_2,\tau_2) = \frac{\prod_{s=1}^{4}\pi_{is}^{\mu_{2s}\tau_2-1}}{\mathrm{D}(\boldsymbol{\mu}_2\tau_2)}. \tag{8}$$

3. For each i, let $\boldsymbol{\psi}_i=(\psi_{i11},\ldots,\psi_{i1c},\psi_{i21},\ldots,\psi_{i2c},\ldots,\psi_{irc})'$ so that $\boldsymbol{\psi}=(\boldsymbol{\psi}'_1,\ldots,\boldsymbol{\psi}'_A)'$. We assume for each i that the ψ_{ijk} are independently and identically

distributed in accordance with a distribution derived from the Gamma(β, β), where the support is confined to the open interval $(0, (1-\pi_{i1})^{-1})$; in other words, the ordinary gamma distribution is truncated as

$$\psi_{ijk} \mid \beta, \boldsymbol{\pi}_i \overset{\text{ind}}{\sim} \text{Gamma}(\beta, \beta) \quad \text{such that } 0 < \psi_{ijk} < (1-\pi_{i1})^{-1}.$$

It is worth noting that these ψ_{ijk} are identically distributed over j and k. Again, one can use other distributions such as a truncated lognormal density, but this will make little difference. In this formulation, there is some information about β because the small areas are assumed to share a common effect.

Thus, for area i, the density for $\boldsymbol{\psi}_i$ is

$$g_{3i}(\boldsymbol{\psi}_i \mid \beta, \boldsymbol{\pi}_i) = \prod_{j=1}^{r}\prod_{k=1}^{c}\left\{ \frac{\beta^\beta \psi_{ijk}^{\beta-1} e^{-\beta\psi_{ijk}}}{\Gamma(\beta)} \Big/ \int_0^{(1-\pi_{i1})^{-1}} \frac{\beta^\beta \psi_{ijk}^{\beta-1} e^{-\beta\psi_{ijk}}}{\Gamma(\beta)} d\psi_{ijk} \right\},$$

For $0 < \psi_{ijk} < (1-\pi_{i1})^{-1}$. Making the transformation $t_{ijk} = \beta\psi_{ijk}$, one can see that the normalizing constant in the denominator of each of the factors in $g_{3i}(\boldsymbol{\psi}_i \mid \beta, \boldsymbol{\pi}_i)$ is $G_\beta[\beta(1-\pi_{i1})^{-1}]$, where $G_\beta(\cdot)$ is the gamma function with scale parameter β. To eliminate the dependence of the range of integration on π_{i1}, let $\phi_{ijk} = (1-\pi_{i1})\psi_{ijk}$ and let $\boldsymbol{\phi}_i = (\phi_{i11}, \ldots, \phi_{i1c}, \phi_{i21}, \ldots, \phi_{i2c}, \ldots, \phi_{irc})'$. Then

$$g_{3i}(\boldsymbol{\phi}_i \mid \beta, \boldsymbol{\pi}_i) = \prod_{j=1}^{r}\prod_{k=1}^{c}\left\{ \frac{\beta^\beta}{(1-\pi_{i1})^\beta} \frac{\phi_{ijk}^{\beta-1} e^{-\frac{\beta\phi_{ijk}}{1-\pi_{i1}}}}{\Gamma(\beta)\, G_\beta[\beta(1-\pi_{i1})^{-1}]} \right\}, \quad (9)$$

for $0 < \phi_{ijk} < 1$. The joint prior for $\boldsymbol{\pi}_i$ and $\boldsymbol{\phi}_i$ is just the product of $g_{3i}(\boldsymbol{\phi}_i \mid \beta, \boldsymbol{\pi}_i)$ and $g_{2i}(\boldsymbol{\pi}_i \mid \boldsymbol{\mu}_2, \tau_2)$. Thus, the joint prior for $\boldsymbol{\phi} = (\boldsymbol{\phi}_1, \ldots, \boldsymbol{\phi}_A)'$ and $\boldsymbol{\pi}$ is

$$g^*(\boldsymbol{\pi}, \boldsymbol{\phi} \mid \boldsymbol{\mu}_2, \tau_2, \beta) \overset{\text{def}}{=} \prod_{i=1}^{A}\{g_{3i}(\boldsymbol{\phi}_i \mid \beta, \boldsymbol{\pi}_i) \cdot g_{2i}(\boldsymbol{\pi}_i \mid \boldsymbol{\mu}_2, \tau_2)\}.$$

That is

$$g^*(\boldsymbol{\pi}, \boldsymbol{\phi} \mid \boldsymbol{\mu}_2, \tau_2, \beta) = \prod_{i=1}^{A}\left\{ \frac{\prod_{s=1}^{4}\pi_{is}^{\mu_{2s}\tau_2 - 1}}{\text{D}(\boldsymbol{\mu}_2\tau_2)} \prod_{j=1}^{r}\prod_{k=1}^{c} \frac{\beta^\beta}{(1-\pi_{i1})^\beta} \frac{\phi_{ijk}^{\beta-1} e^{-\frac{\beta\phi_{ijk}}{1-\pi_{i1}}}}{\Gamma(\beta)\, G_\beta[\beta(1-\pi_{i1})^{-1}]} \right\}. \quad (10)$$

The description of the model is completed by specifying the assumptions on the hyperparameters. As there are no conjugate priors, we use shrinkage priors for τ_1, τ_2 and β because these are proper and noninformative. Priors of the form $p(\tau_1) \propto 1/\tau_1$, and specifically proper diffused gamma priors, are discouraged; see, for example, Gelman (2006). Other alternatives are half Cauchy densities and gamma densities (one would need to specify the hyper-parameters). Thus, we take

1. τ_1, τ_2 and β have independent shrinkage priors of the form

$$f(x) = \frac{a_0}{(a_0 + x)^2}, \text{ for } x \geq 0,$$

where a_0 is specified; it is standard practice to take $a_0 = 1$.

2. We also assume that $\boldsymbol{\mu}_1 \sim \text{Dirichlet}(1, 1, \ldots, 1)$ and $\boldsymbol{\mu}_2 \sim \text{Dirichlet}(1, 1, 1, 1)$.

Let $\Omega = (\beta, \boldsymbol{\mu}_1, \tau_1, \boldsymbol{\mu}_2, \tau_2)$. Then the density for Ω is

$$p(\Omega) = \frac{a_0}{(a_0+\tau_1)^2} \cdot \frac{b_0}{(b_0+\tau_2)^2} \cdot \frac{c_0}{(c_0+\beta)^2} (rc-1)!3!$$

for τ_1, τ_2 and $\beta \geq 0$, $\sum_{j,k} \mu_{1jk} = 1$ and $\sum_{s=1}^{4} \mu_{2s} = 1$.

By Bayes' theorem, the joint posterior density is

$$\begin{aligned} h(\Omega, \boldsymbol{p}, \boldsymbol{\pi}, \boldsymbol{\phi}, \boldsymbol{y}_{(1)} \mid \boldsymbol{y}_1, \boldsymbol{u}, \boldsymbol{v}, \boldsymbol{w}) \propto & \\ & f(\boldsymbol{y} \mid \boldsymbol{\pi}, \boldsymbol{\phi}, \boldsymbol{p})\, g_1(\boldsymbol{p} \mid \boldsymbol{\mu}_1, \tau_1)\, g^*(\boldsymbol{\pi}, \boldsymbol{\phi} \mid \boldsymbol{\mu}_2, \tau_2, \beta)\, p(\Omega) \\ = & \prod_{i=1}^{A}\Bigg[\binom{n_i}{\boldsymbol{y}'_{i1}, \boldsymbol{y}'_{i2}, \boldsymbol{y}'_{i3}, \boldsymbol{y}'_{i4}} (1-\pi_{i1})^{-y_{i..}} \\ & \left\{ \prod_{s=1}^{4} \pi_{is}^{y_{is.}} \times \prod_{j,k} \{p_{ijk}\phi_{ijk}\}^{y_{i.jk}} \left[a_{ijk}(\pi_{i1}, \phi_{ijk}) \right]^{y_{i1jk}} \right\} \\ & \times \left\{ \frac{\prod_{j,k} p_{ijk}^{\mu_{1jk}\tau_1 - 1}}{\text{D}(\boldsymbol{\mu}_1\tau_1)} \right\} \left\{ \frac{\prod_{s=1}^{4} \pi_{is}^{\mu_{2s}\tau_2 - 1}}{\text{D}(\boldsymbol{\mu}_2\tau_2)} \right. \\ & \left. \left. \prod_{j=1}^{r}\prod_{k=1}^{c} \frac{\beta^\beta}{(1-\pi_{i1})^\beta} \frac{\phi_{ijk}^{\beta-1} e^{-\frac{\beta\phi_{ijk}}{1-\pi_{i1}}}}{\Gamma(\beta)\, G_\beta[\beta(1-\pi_{i1})^{-1}]} \right\} \right] \\ & \times \frac{a_0}{(a_0+\tau_1)^2} \cdot \frac{b_0}{(b_0+\tau_2)^2} \cdot \frac{c_0}{(c_0+\beta)^2}, \end{aligned} \quad (11)$$

where, substituting $(1-\pi_{i1})^{-1}\phi_{ijk}$ for ψ_{ijk},

$$a_{ijk}(\pi_{i1}, \phi_{ijk}) = \left(\frac{1-\pi_{i1}}{\phi_{ijk}} \right) \left[1 + \frac{1}{\pi_{i1}} \{1 - \pi_{i1} - \phi_{ijk}\} \right]. \quad (12)$$

To make inferences about the p_{ijk}, we will draw samples from $h(\Omega, \boldsymbol{p}, \boldsymbol{\pi}, \boldsymbol{\phi}, \boldsymbol{y}_{(1)} \mid \boldsymbol{y}_1, \boldsymbol{u}, \boldsymbol{v}, \boldsymbol{w})$ using Markov chain Monte Carlo methods. This procedure is described in Section 3.

3. Computations

We use the SIR algorithm to subsample a random sample from an approximate posterior density. There are three steps to accomplish this task. We collapse over the $\boldsymbol{p}_i, \boldsymbol{\pi}_i$ and $\boldsymbol{\phi}_i$, approximate the collapsed density by a simpler one and sample from it, and then subsample these samples to get samples from the original density. We show how to do these three steps in this section.

To obtain the approximation and to simplify the computations, in Appendix A we collapse over the $\boldsymbol{p}_i, \boldsymbol{\pi}_i$ and $\boldsymbol{\phi}_i$ to get

$$h(\Omega, \boldsymbol{y}_{(1)} \mid \boldsymbol{y}_1, \boldsymbol{u}, \boldsymbol{v}, \boldsymbol{w}) = \pi_a(\Omega\, \boldsymbol{y}_{(1)} \mid \boldsymbol{y}_1, \boldsymbol{u}, \boldsymbol{v}, \boldsymbol{w}) \cdot \prod_{i=1}^{A} I_i,$$

where

$$I_i = \int\int\int_0^{\frac{\beta b_i}{1-\pi_{i1}}} \frac{G_{rc\beta}\left(\frac{\beta b_i}{1-\pi_{i1}}\right)}{\left[G_\beta\left(\frac{\beta}{1-\pi_{i1}}\right)\right]^{rc}}$$
$$\prod_{j,k}\left\{\left(\frac{W_i}{\beta}\right)\sum_{s=2}^{4} y_{isjk}\left[\frac{1}{\phi^*_{ijk}}\left(1+\frac{1-\pi_{i1}}{\pi_{i1}}\left\{1-\frac{W_i\phi^*_{ijk}}{\beta}\right\}\right)\right]^{y_{i1\,jk}}\right\}$$
$$\frac{W_i^{rc\beta-1}\, e^{-W_i}}{\Gamma(rc\beta)G_{rc\beta}\left(\frac{\beta b_i}{1-\pi_{i1}}\right)}\, \mathrm{d}W_i \left\{\frac{\prod_{j,k}\phi^{*\, y_{i\cdot jk}+\beta-1}_{ijk}}{\mathrm{D}(\boldsymbol{y}_i^{(1)}+\beta\boldsymbol{j})}\right\}$$
$$\left\{\frac{\prod_{s=1}^{4}\pi_{is}^{y_{is\cdot}+\mu_2 s\tau_2-1}}{\mathrm{D}(\boldsymbol{y}_i^{(2)}+\boldsymbol{\mu}_2\tau_2)}\right\}\mathrm{d}\boldsymbol{\phi}^*_i\,\mathrm{d}\boldsymbol{\pi}_i, \qquad (13)$$

with $b_i = \min\{\{1/\phi^*_{ijk}\},\ j=1,\ldots,r;\ k=1,\ldots,c\}$ and

$$\pi_a(\Omega, \boldsymbol{y}_{(1)} \mid \boldsymbol{y}_1, \boldsymbol{u}, \boldsymbol{v}, \boldsymbol{w})$$
$$= \frac{a_0}{(a_0+\tau_1)^2}\cdot\frac{b_0}{(b_0+\tau_2)^2}\cdot\frac{c_0}{(c_0+\beta)^2}$$
$$\prod_{i=1}^{A}\frac{\Gamma(rc\beta)}{[\Gamma(\beta)]^{rc}}\mathrm{D}(\boldsymbol{y}_i^{(1)}+\beta\boldsymbol{j})$$
$$\times\prod_{i=1}^{A}\binom{n_i}{\boldsymbol{y}'_{i1},\boldsymbol{y}'_{i2},\boldsymbol{y}'_{i3},\boldsymbol{y}'_{i4}}\frac{\mathrm{D}(\boldsymbol{y}_i^{(1)}+\boldsymbol{\mu}_1\tau_1)}{\mathrm{D}(\boldsymbol{\mu}_1\tau_1)}\frac{\mathrm{D}(\boldsymbol{y}_i^{(2)}+\boldsymbol{\mu}_2\tau_2)}{\mathrm{D}(\boldsymbol{\mu}_2\tau_2)}. \qquad (14)$$

To evaluate I_i for each $i=1,\ldots,A$, we proceed as follows given $(\Omega, \boldsymbol{y}_{i(1)})$:

1. Draw independent samples of vectors $\boldsymbol{\pi}_i$ and $\boldsymbol{\phi}^*_i$ from the Dirichlet$(\boldsymbol{y}_i^{(2)}+\boldsymbol{\mu}_2\tau_2)$ and Dirichlet$(\boldsymbol{y}_i^{(1)}+\beta\,\boldsymbol{j})$, respectively. For each $\boldsymbol{\pi}_i$ and $\boldsymbol{\phi}^*_i$, draw a sample of values for W_i from the truncated gamma distribution on the interval $(0, \{\beta b_i/1-\pi_{i1}\})$ with parameter $rc\beta$.
2. For each $\boldsymbol{\pi}_i, \boldsymbol{\phi}^*_i$ and W_i selected in step (1), compute R_1R_2, where

$$R_1 = G_{rc\beta}\left(\frac{\beta b_i}{1-\pi_{i1}}\right)\Bigg/\left[G_\beta\left(\frac{\beta}{1-\pi_{i1}}\right)\right]^{rc} \qquad (15)$$

and

$$R_2 = \prod_{j,k}\left\{\left(\frac{W_i}{\beta}\right)\sum_{s=2}^{4} y_{isjk}\left[\frac{1}{\phi^*_{ijk}}\left(1+\frac{1-\pi_{i1}}{\pi_{i1}}\left\{1-\frac{W_i\phi^*_{ijk}}{\beta}\right\}\right)\right]^{y_{i1\,jk}}\right\}. \qquad (16)$$

3. Repeat steps (1) and (2) 1,000 times. Then compute the average of R_1R_2 over these 1,000 values.

The rest of our computation has two parts. First, we use the griddy Metropolis-Hastings sampler to draw from $\pi_a(\Omega, \boldsymbol{y}_{(1)} \mid \boldsymbol{y}_1, \boldsymbol{u}, \boldsymbol{v}, \boldsymbol{w})$. We sample μ_1, μ_2, τ_1 and τ_2 from their conditional posterior densities using grids; this entails transforming τ_1 and τ_2 to the unit interval $(0, 1)$. For each distribution, 100 grids are used; see Nandram, Cox and Choi (2005) for a similar procedure. Here $\boldsymbol{y}_{(1)}$ is drawn by sampling from its conditional probability mass function component wise. Draws are made from the conditional posterior density of β using a Metropolis step in a manner similar to Nandram and Choi (2002a, b). We have performed this algorithm 11,000 times and we allowed a "burn-in" of 1,000 iterates. We found that the autocorrelations among the iterates was small, thereby indicating strong mixing of the sampler. We have also used the batch-means method to further assess the computation. We used batches of 25 to compute numerical standard errors.

Second, we use the SIR algorithm to subsample the sample of 10,000 iterates we obtained from $\pi_a(\Omega, \boldsymbol{y}_{(1)} \mid \boldsymbol{y}_1, \boldsymbol{u}, \boldsymbol{v}, \boldsymbol{w})$. For each of the 10,000 iterates we calculate the weights

$$w_m = \frac{h(\Omega^{(m)}, \boldsymbol{y}_{(1)}^{(m)} \mid \boldsymbol{y}_1, \boldsymbol{u}, \boldsymbol{v}, \boldsymbol{w})}{\pi_a(\Omega^{(m)}, \boldsymbol{y}_{(1)}^{(m)} \mid \boldsymbol{y}_1, \boldsymbol{u}, \boldsymbol{v}, \boldsymbol{w})},$$
$$m = 1, \ldots, M = 10{,}000, \qquad (17)$$

and we resample $\{\Omega^{(m)}, \boldsymbol{y}_{(1)}^{(m)}\}$ with probabilities proportional to the weights w_m for $m=1,\ldots,M$ without replacement. We use a 10% sampling, and we subsample the 10,000 iterates to get 1,000 iterates; sampling without replacement is a good idea because it avoids repeated values which already exist because the Metropolis-Hastings sampler is not really an accept-reject sampler and it gives repeated values. As usual with sampling without replacement the weights are calculated every time a value is selected.

Finally, we can now make exact (within limitations of Markov chain Monte Carlo methods) inference about $\boldsymbol{p}_i$ a posteriori. Letting $y_{i\cdot jk} = \sum_{s=1}^{4} y_{isjk}$ and $\boldsymbol{y}_i^*$ denote the vector of $y_{i\cdot jk}$. Then,

$$\boldsymbol{p}_i \mid \boldsymbol{y}_i^*, \boldsymbol{\mu}_1, \tau_1 \overset{\text{ind}}{\sim} \text{Dirichlet}(\boldsymbol{y}_i^* + \boldsymbol{\mu}_1\tau_1),\ i=1,\ldots,A.$$

Thus, for each value of $\boldsymbol{y}_i^*, \boldsymbol{\mu}_1$, and τ_i we obtain from the SIR algorithm, we draw a value of $\boldsymbol{p}_i, i = 1, \ldots, A$. Thus, we obtain a Rao-Blackwellized density for each of the $\boldsymbol{p}_i$, and inference proceeds in the usual way.

4. An illustrative example

Our illustrative example is in health statistics. In Section 4.1 we briefly discuss the data we used from the third National Health and Nutrition Examination Survey (NHANES III). Specifically, we study the relationship between bone mineral density and family income; see Nandram, Cox and Choi (2005) for a discussion of this problem. In Section 4.2, after briefly discussing our computation, we present posterior inference on the cell probabilities. In section 4.3 using the Bayes factor we discuss the relation between BMD and FI.

4.1 NHANES III data

The sample design is a stratified multistage probability design which is representative of the total civilian non-institutionalized population, 2 months of age or older, in the United States. Further details of the NHANES III sample design are available (National Center for Health Statistics 1992, 1994). The NHANES III data collection consists of two parts: the first part is the sample selection and the interview of the members of a sampled household for their personal information, and the second part is the examination of those interviewed at the mobile examination center (MEC). The health examination has information on physical examination, tests and measurements performed by technicians, and specimen collection. The sample was selected from households in 81 primary units across the continental United States during the period from October 1988 through September 1994. The final data for this study is a part of the 35 largest primary sampling units with population at least 500,000, and we consider 13 subnational areas.

Nonresponse occurs in the interview and examination parts of the survey. The interview nonresponse arises from sampled persons who did not respond for the interview. Some of those who were already interviewed and included in the subsample for a health examination missed the examination at home or at the MEC, thereby missing all or part of the examinations.

Doctors believe that obese and overweight individuals do not generally turn up at the MEC. Cohen and Duffy (2002) point out that "Health surveys are a good example, where it seems plausible that propensity to respond may be related to health." NHANES III is a good example.

Sampled persons in NHANES III can be categorized by many types of attributes, and researchers analyze such categorical tables for goodness of fit or independence. Here we study bone mineral density (BMD) and family income (FI). We note here that while FI is a discrete variable, we have classified BMD into three levels (normal, osteopenia and osteoporosis), and FI into three levels (low, medium and high). However, only partial classification of the individuals is available because some individuals are classified by only one attribute while others are not classified. About 62% of the households have both FI and BMD observed, 8% with only BMD observed, 29% with only income observed, 1% with neither income nor BMD among those participated in the examination stage. Our problem is to estimate the cell probabilities and to test for association between BMD and FI for each of 13 subnational areas using our expansion model that pools the data adaptively.

In Table 1 we present the 3×3 tables of BMD and FI for the aforementioned 13 areas. Note that areas 6 and 48 have enough data so that they can stand by themselves. However, the other areas are very small; the counts in the table with row totals are generally small except for area 17 and the counts in the table with just total are small. Even for the table with complete data the cell counts are generally small forcing us to use small area estimation techniques to borrow strength.

Table 1
Counts of the 3×3 tables of BMD and FI corresponding to 13 subnational areas in NHANES III

State	Complete Table									Column Total			Row Total			Total
4	21	14	9	8	7	3	2	2	0	11	5	6	4	0	1	1
6	257	127	106	92	51	32	32	5	7	178	54	82	65	28	4	20
12	33	18	21	22	4	4	15	5	0	18	11	16	5	6	2	1
17	25	15	13	8	5	3	0	0	1	18	10	16	17	2	2	4
25	9	7	12	6	5	9	2	1	0	9	6	12	1	4	5	1
26	18	11	18	6	5	9	2	1	1	10	5	11	4	3	0	1
29	9	4	10	3	2	4	3	1	2	9	2	9	0	2	4	1
36	42	17	27	32	13	18	9	6	1	43	21	42	9	7	6	1
39	8	6	14	2	5	4	3	0	1	9	7	5	2	3	0	0
42	14	8	11	12	8	4	8	1	2	35	15	24	3	1	0	0
44	12	9	6	8	5	0	5	1	0	19	4	12	7	1	0	1
48	159	44	22	51	11	13	9	6	2	88	12	23	16	8	2	14
53	14	10	15	10	10	14	3	1	1	9	4	8	2	4	1	0

Note: In the complete 3×3 table the first (second, third) set of three numbers is the first (second, third) row; the column (row) total refers to the 3×3 table with only column (row) totals; the total refers to the 3×3 table with only total.

4.2 Posterior inference of the cell probabilities

We discuss the performance of our computations for the expansion model, and then we discuss posterior inference about the cell probabilities. We use the posterior mean (PM), posterior standard deviation (PSD) and 95% credible interval for each of parameters of interest. We also present the numerical standard errors (NSE) to assess the repeatability of our computations.

In Table 2 we present summaries of the posterior distributions of $\boldsymbol{\mu}_1, \boldsymbol{\mu}_2, \tau_1, \tau_2$ and β, both before and after the application of the SIR algorithm. These summaries are very similar indicating that the SIR approximation $\pi_a(\Omega, \boldsymbol{y}_{(1)} \mid \boldsymbol{y}_1, \boldsymbol{u}, \boldsymbol{v}, \boldsymbol{w})$ is not unreasonable. For example, a 95% credible interval for β before and after the application of the SIR algorithm are respectively (1.081, 1.940) and (1.086, 1.947), very good agreement. The estimates of τ_1 and τ_2 should show the largest discrepancies, but these are also reasonably close [*e.g.*, for τ_1 95% credible intervals with the approximation is (28.282, 64.204) and with the SIR algorithm it is (27.962, 64.425)]. In both cases the NSEs are small indicating that the computations are repeatable.

In Table 3 we have compared our expansion model (Model 3) with two other models. Model 1, an ignorable nonresponse model, and Model 2, a nonignorable non-response model (no centering), are described in Appendix B. For illustration we have selected three areas, a large area, a medium-sized area and a small area. There are differences among the three models. In general, the larger estimates tend to be smaller for Model 2, and even smaller than Model 1, than for Model 3 (*i.e.*, the estimates from Model 3 are naturally closest to Model 1, and not Model 2). Model 2 produces the largest variability; as expected, Model 3 gives slightly larger variability than Model 1. Because of space restrictions we have not presented the NSEs, but we note that they are all smaller than 0.005.

4.3 Bayes factor for evidence of association

We have also considered the association between BMD and FI. It appears doubtful whether such an association might exist, but it is interesting to look at this issue; see Nandram, Cox and Choi (2005) for a discussion on this problem. We use the Bayes factor (Kass and Raftery 1995) to measure the strength of the evidence of an association relative to no association in the $r \times c$ categorical table. We have done so for each of the thirteen areas and all areas combined.

We have used two procedures, one without extensive modeling and the other using our nonignorable nonresponse (expansion) model. The simple method is to fill in the cell counts using an ordinary raking procedure, and we assume there is no error in doing so. This is a common sense procedure that survey practitioners have used routinely. In the second procedure using our nonignorable nonresponse model, we have obtained 1,000 combined tables for each area as described in Section 3 on computations. For each area we have obtained the cell counts for all four tables, and we summed them to get a single table of all counts.

Table 2
NHANES data on 13 areas: Comparison of the approximate posterior density and the correct posterior density using the posterior means (PM), posterior standard deviations (PSD), numerical standard errors (NSE) and 95% credible intervals of the hyperparameters

	Approximation				Adjusted			
	PM	**PSD**	**NSE**	**95% Int**	**PM**	**PSD**	**NSE**	**95% Int**
μ_{21}	0.528	0.031	0.001	(0.463, 0.582)	0.525	0.031	0.008	(0.456, 0.578)
μ_{22}	0.131	0.021	0.001	(0.096, 0.181)	0.133	0.021	0.002	(0.094, 0.179)
μ_{23}	0.328	0.028	0.001	(0.274, 0.383)	0.328	0.028	0.005	(0.269, 0.383)
μ_{24}	0.013	0.006	0.000	(0.004, 0.027)	0.014	0.006	0.000	(0.004, 0.029)
τ_2	21.638	9.559	0.255	(8.347, 46.587)	20.078	8.632	0.303	(8.538, 38.625)
μ_{111}	0.280	0.023	0.001	(0.234, 0.324)	0.277	0.023	0.004	(0.228, 0.319)
μ_{112}	0.133	0.016	0.000	(0.102, 0.165)	0.134	0.017	0.002	(0.101, 0.165)
μ_{113}	0.200	0.019	0.000	(0.163, 0.238)	0.199	0.019	0.003	(0.162, 0.236)
μ_{121}	0.105	0.015	0.000	(0.078, 0.135)	0.107	0.015	0.002	(0.079, 0.135)
μ_{122}	0.065	0.011	0.000	(0.044, 0.088)	0.065	0.011	0.001	(0.044, 0.087)
μ_{123}	0.072	0.012	0.000	(0.050, 0.096)	0.073	0.012	0.001	(0.049, 0.097)
μ_{131}	0.061	0.011	0.000	(0.041, 0.083)	0.061	0.011	0.001	(0.040, 0.083)
μ_{132}	0.037	0.008	0.000	(0.023, 0.054)	0.036	0.008	0.001	(0.022, 0.054)
μ_{133}	0.048	0.009	0.000	(0.031, 0.068)	0.048	0.009	0.001	(0.031, 0.068)
τ_1	45.960	10.094	0.153	(28.282, 64.204)	45.177	10.562	0.679	(27.962, 64.423)
β	1.472	0.218	0.004	(1.081, 1.940)	1.449	0.208	0.022	(1.086, 1.947)

Note: The hyperparameters are $\boldsymbol{\mu}_1, \boldsymbol{\mu}_2, \tau_1, \tau_2$ and β.

Table 3
Posterior means of the cell probabilities and 95% credible intervals (CI) for three areas (large, medium and small) by the three models

	Model 1			Model 2			Model 3		
Cell	**PM**	**PSD**	**95% CI**	**PM**	**PSD**	**95% CI**	**PM**	**PSD**	**95% CI**
a. Large									
(1,1)	0.239	0.044	(0.157, 0.326)	0.196	0.046	(0.117, 0.295)	0.259	0.038	(0.189, 0.335)
(1,2)	0.140	0.035	(0.078, 0.213)	0.127	0.035	(0.068, 0.200)	0.132	0.029	(0.082, 0.197)
(1,3)	0.240	0.044	(0.159, 0.332)	0.198	0.047	(0.118, 0.301)	0.248	0.037	(0.175, 0.322)
(2,1)	0.092	0.032	(0.039, 0.162)	0.098	0.040	(0.037, 0.188)	0.077	0.022	(0.039, 0.126)
(2,2)	0.074	0.028	(0.029, 0.136)	0.077	0.030	(0.030, 0.144)	0.056	0.020	(0.024, 0.099)
(2,3)	0.133	0.036	(0.070, 0.210)	0.121	0.042	(0.056, 0.219)	0.110	0.028	(0.058, 0.168)
(3,1)	0.036	0.020	(0.008, 0.083)	0.069	0.039	(0.013, 0.153)	0.047	0.018	(0.018, 0.086)
(3,2)	0.023	0.015	(0.003, 0.061)	0.043	0.025	(0.007, 0.100)	0.032	0.014	(0.009, 0.063)
(3,3)	0.025	0.017	(0.003, 0.066)	0.071	0.040	(0.010, 0.154)	0.042	0.016	(0.016, 0.079)
b. Medium									
(1,1)	0.233	0.034	(0.169, 0.302)	0.213	0.043	(0.141, 0.305)	0.254	0.032	(0.194, 0.318)
(1,2)	0.143	0.028	(0.093, 0.200)	0.127	0.032	(0.072, 0.196)	0.146	0.024	(0.102, 0.197)
(1,3)	0.190	0.031	(0.132, 0.254)	0.140	0.034	(0.084, 0.218)	0.208	0.027	(0.156, 0.259)
(2,1)	0.174	0.031	(0.118, 0.237)	0.160	0.042	(0.092, 0.249)	0.154	0.027	(0.106, 0.211)
(2,2)	0.043	0.018	(0.015, 0.083)	0.060	0.028	(0.017, 0.124)	0.032	0.012	(0.012, 0.059)
(2,3)	0.049	0.020	(0.017, 0.095)	0.065	0.031	(0.018, 0.136)	0.042	0.014	(0.020, 0.072)
(3,1)	0.112	0.025	(0.068, 0.167)	0.120	0.041	(0.059, 0.209)	0.092	0.020	(0.056, 0.134)
(3,2)	0.047	0.018	(0.018, 0.088)	0.059	0.026	(0.019, 0.118)	0.040	0.014	(0.018, 0.071)
(3,3)	0.010	0.009	(0.000, 0.033)	0.056	0.032	(0.006, 0.122)	0.032	0.012	(0.013, 0.059)
c. Small									
(1,1)	0.196	0.052	(0.103, 0.305)	0.164	0.055	(0.077, 0.288)	0.253	0.043	(0.175, 0.334)
(1,2)	0.081	0.034	(0.028, 0.158)	0.081	0.032	(0.030, 0.155)	0.091	0.028	(0.043, 0.152)
(1,3)	0.213	0.052	(0.118, 0.323)	0.175	0.055	(0.087, 0.300)	0.220	0.043	(0.137, 0.306)
(2,1)	0.093	0.041	(0.028, 0.186)	0.111	0.055	(0.029, 0.234)	0.073	0.028	(0.030, 0.139)
(2,2)	0.056	0.029	(0.012, 0.126)	0.066	0.031	(0.018, 0.136)	0.045	0.020	(0.014, 0.094)
(2,3)	0.115	0.045	(0.042, 0.215)	0.118	0.053	(0.038, 0.240)	0.092	0.030	(0.041, 0.158)
(3,1)	0.115	0.048	(0.036, 0.222)	0.113	0.056	(0.031, 0.239)	0.081	0.030	(0.033, 0.148)
(3,2)	0.044	0.028	(0.006, 0.113)	0.065	0.035	(0.013, 0.144)	0.043	0.020	(0.012, 0.086)
(3,3)	0.087	0.042	(0.022, 0.184)	0.107	0.055	(0.023, 0.227)	0.103	0.034	(0.047, 0.181)

Note: See Appendix B for a description of Models 1 and 2.

The raking procedure to obtain the cell counts is described as follows. Let n_{jk} denote the cell counts for the four tables combined. Let $n_{jk}^{(1)}$ denote the cell counts for the table with complete data, $n_{j,c+1}^{(2)}$ denote the table with row totals, $n_{r+1,k}^{(3)}$ denote the table with column totals and $n_{r+1,c+1}^{(4)}$ denote the table with total. The cell counts for the four tables are estimated as

$$n_{jk} = n_{jk}^{(1)} + \left(\frac{n_{jk}^{(1)}}{n_{j\cdot}^{(1)}}\right) n_{j,c+1}^{(2)} + \left(\frac{n_{jk}^{(1)}}{n_{k}^{(1)}}\right) n_{r+1,k}^{(3)} + \left(\frac{n_{jk}^{(1)}}{n_{\cdot\cdot}^{(1)}}\right) n_{r+1,c+1}^{(4)},$$

$j = 1,\ldots,r, k = 1,\ldots,c.$

In either case we denote the sum of the cell counts for each area by n_{jk}. For the raking procedure we have a single table for each area, and for the nonignorable nonresponse model we have a sample of 1,000 tables for each area. We also have a single combined table for all areas under the raking procedure and 1,000 tables for all areas combined. We obtain the Bayes factor for each table under a multinomial-Dirichlet model. It is worth noting that our method uses the expansion model so that the cell counts borrow strength from other areas unlike the raking procedure.

Then, for each table we take

$$\boldsymbol{n} \mid \boldsymbol{\pi} \sim \text{Multinomial}(n, \boldsymbol{\pi}) \text{ and } \boldsymbol{\pi} \sim \text{Dirichlet}(\mathbf{1}).$$

That is, we take a uniform prior for $\boldsymbol{\pi}$ with $\pi_{jk} > 0$ and $\sum_{j=1}^{r}\sum_{k=1}^{c}\pi_{jk} = 1$. Under the hypothesis of no association we have $\pi_{jk} = \pi_j \pi_k$, where $\pi_j > 0, \sum_{j=1}^{r}\pi_j = 1$ and $\pi_k > 0, \sum_{k=1}^{c}\pi_k = 1$. Thus, the hypothesis of association is that the π_{jk} are unrestricted (except that they are nonnegative and sum to unity) whereas for the hypothesis of no associate $\pi_{jk} = \pi_j \pi_k$.

The Bayes factor is the ratio of the marginal likelihood under association versus the marginal likelihood under no association. This measures the strength of evidence of association versus no association; see Kass and Raftery (1995). Let $p_a(\boldsymbol{n})$ denote the marginal likelihood under association and $p_0(\boldsymbol{n})$ denote the marginal likelihood under no association. Then, letting $n_{j\cdot} = \sum_{k=1}^{c} n_{jk}$ and $n_{\cdot k} = \sum_{j=1}^{r} n_{jk}$, it is easy to show that

$$p_a(\boldsymbol{n}) = p_0(\boldsymbol{n})\left\{\prod_{u=0}^{n-1}\frac{u+rc}{(u+r)(u+c)}\right\}\frac{\prod_{j=1}^{r} n_{j\cdot}!\prod_{k=1}^{c} n_{\cdot k}!}{\prod_{j=1}^{r}\prod_{k=1}^{c} n_{jk}!},$$

where $p_0(\boldsymbol{n}) = n!\prod_{u=0}^{n-1}(u+rc)^{-1}$. Observe that $p_0(\boldsymbol{n})$ is not a function of $\{n_{jk}\}$. Thus, as a measure of association it is the deviation of $\prod_{j=1}^{r} n_{j\cdot}!\prod_{k=1}^{c} n_{\cdot k}!$ from $\prod_{j=1}^{r}\prod_{k=1}^{c} n_{jk}!$ that matters. However, we note that for the classical Pearson statistic for independence it is the deviations of n_{jk} from $n_{j\cdot}n_{\cdot k}$ that matter. But note that this test cannot be applied because many of the expected cell counts are smaller than 5 under the hypothesis of no association and multinomial sampling.

We present our results in Table 4 and in Figure 1 corresponding to the data in Table 1 for the cross-classification of BMD and FI. We have presented the logarithms of the marginal likelihoods (base *e*) and the Bayes factors; these are to be interpreted using the rule of thumb of Kass and Raftery (1995).

In Figure 1 we can see that the box plots are all above zero except the one for the third area which provides no evidence for association; perhaps there is no evidence for association in area 42 (10 in figure) as well. A summary of these results are presented in Table 4. The Bayes factors show association in all areas, except area 12, and they are much larger under the nonignorable nonresponse model. Area 6 and all areas combined are elevated (336.3 *vs.* 5.8 and 3,798.2 *vs.* 0.183).

Table 4
NHANES data on 13 areas: Comparison of the negative marginal likelihoods and Bayes factors or association of BMD and F1 from the raking procedure and the expansion model by area

	Raking			Expansion	
area	$-\ln\{p_0(n)\}$	$-\ln\{p_a(n)\}$	*BF*	$-\ln\{p_a(n)\}$	*BF*
4	26.19	23.07	22.855	$23.5_{0.014}$	$14.78_{0.169}$
6	45.73	43.98	5.766	$40.5_{0.038}$	$336.27_{11.465}$
12	31.14	38.01	0.001	$33.4_{0.054}$	$0.37_{0.027}$
17	29.13	27.03	8.134	$27.0_{0.026}$	$10.27_{0.191}$
25	25.44	26.02	0.558	$23.8_{0.029}$	$9.55_{0.202}$
26	26.89	23.18	40.562	$23.9_{0.018}$	$24.71_{0.370}$
29	23.21	20.87	10.301	$21.3_{0.018}$	$8.40_{0.115}$
36	34.99	36.09	0.330	$33.1_{0.064}$	$21.13_{0.928}$
39	23.77	24.89	0.325	$23.6_{0.044}$	$2.24_{0.68}$
42	29.51	30.21	0.497	$30.3_{0.099}$	$4.33_{0.255}$
44	25.61	30.48	0.008	$24.4_{0.027}$	$5.19_{0.137}$
48	38.83	35.34	32.650	$39.1_{0.060}$	$2.15_{0.081}$
53	27.11	24.82	9.865	$24.2_{0.017}$	$19.40_{0.282}$
All	53.43	55.13	0.183	$46.1_{0.049}$	$3{,}798.24_{151.82}$

Note: Area 'all' refers to all areas combined; the notation a_b means that the average is *a* and the standard error is *b* over the 1,000 iterates; $\ln\{p_0(n)\}$ is the same for both procedures.

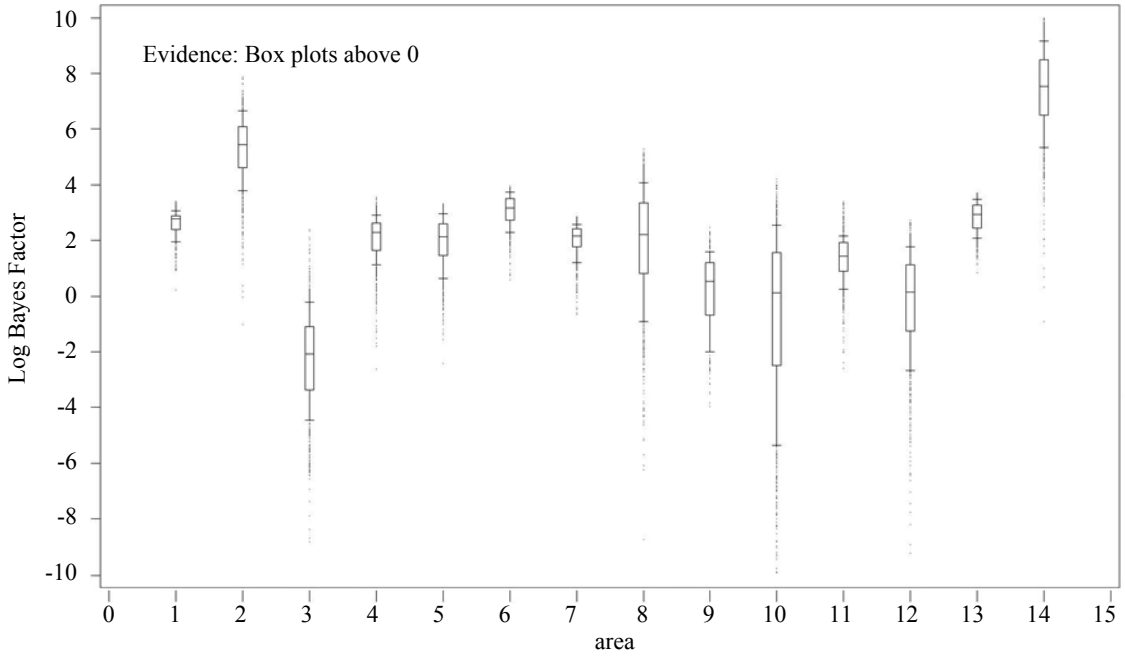

Figure 1 Box plots of log Bayes factors versus areas to measure evidence for association between BMD and FI

5. Concluding remarks

The purpose of this paper has been to develop a methodology to analyze data from incomplete two-way categorical tables, each table corresponding to an area. We have done so by extending the Bayesian methodology of Nandram and Choi (2002a, b) for binary data to $r \times c$ categorical tables for small areas. We have constructed a new Bayesian nonignorable nonresponse model (*i.e.*, expansion model) which is centered on the ignorable nonresponse model. We have used Markov chain Monte Carlo methods (specifically the griddy Metropolis-Hastings sampler) to fit the model. We have compared our model with an ignorable nonresponse model and a nonignorable nonresponse model. Finally, we have done an illustrative example on estimating the cell probabilities for the 3×3 table of BMD and income over thirteen subnational areas.

We have shown that there are differences among the three models. Using the data on BMD and FI, we have shown that our expansion model is a compromise between the ignorable nonresponse model and the nonignorable nonresponse model. Using the Bayes factor we have shown that there are differences between the tests of association for BMI and FI when the cell counts are estimated from our model and when using a raking procedure. In fact, owing to the borrowing of strength, we have seen that the evidence for association under our model is much stronger than the from the raking procedure.

There are three additional avenues that we can explore. First, we can construct a model to incorporate systematic departure from ignorability. This task will need more costly field work to get the much-needed information. Second, it is also interesting to relax the assumption that the margins of the categorical table are fixed; see, for example, Nandram (2009) who looked at a single large area. Third, there can be further improvement in calibration (*i.e.*, incorporating information about margins).

Acknowledgements

We are grateful to the two referees and the associate editor for their assistance in the presentation of the material.

Appendix A

Joint posterior density of the expansion model

First, integrating the joint posterior density over $\boldsymbol{p}$ we get that

$$h(\Omega,\boldsymbol{\pi},\boldsymbol{\phi},\boldsymbol{y}_{(1)}|\boldsymbol{y}_1,\boldsymbol{u},\boldsymbol{v},\boldsymbol{w}) \propto \prod_{i=1}^{4}\frac{\mathrm{D}(\boldsymbol{y}_i^{(1)}+\boldsymbol{\mu}_1\tau_1)}{\mathrm{D}(\boldsymbol{\mu}_1\tau_1)}$$
$$\left[\binom{n_i}{\boldsymbol{y}'_{i1},\,\boldsymbol{y}'_{i2},\,\boldsymbol{y}'_{i3},\,\boldsymbol{y}'_{i4}}(1-\pi_{i1})^{-y_{i\cdot\cdot\cdot}}\right.$$
$$\times\left\{\prod_{s=1}^{4}\pi_{is}^{y_{is\cdot\cdot}}\prod_{j,k}\phi_{ijk}^{y_{i\cdot jk}}[a_{ijk}(\pi_{i1},\phi_{ijk})]^{y_{i1jk}}\right\}$$
$$\times\left\{\frac{\prod_{s=1}^{4}\pi_{is}^{\mu_{2s}\tau_2-1}}{\mathrm{D}(\boldsymbol{\mu}_2\tau_2)}\prod_{j=1}^{r}\prod_{k=1}^{c}\frac{\beta^{\beta}}{(1-\pi_{i1})^{\beta}}\right.$$
$$\left.\left.\frac{\phi_{ijk}^{\beta-1}e^{-\frac{\beta\phi_{ijk}}{1-\pi_{i1}}}}{\Gamma(\beta)G_{\beta}[\beta(1-\pi_{i1})^{-1}]}\right\}\right]$$
$$\times\frac{a_0}{(a_0+\tau_1)^2}\cdot\frac{b_0}{(b_0+\tau_2)^2}\cdot\frac{c_0}{(c_0+\beta)^2}, \qquad (A.1)$$

where the $rc\times 1$ vector

$$\boldsymbol{y}_i^{(1)} \overset{\text{def}}{=} (y_{i.11}, y_{i.12}, \ldots, y_{i.1c}, y_{i.21}, \ldots, y_{i.2c}, y_{i.r1}, \ldots, y_{i.rc})'.$$

Now let $\boldsymbol{j}$ denote an $rc\times 1$ vector of ones and let

$$\boldsymbol{y}_i^{(2)} \overset{\text{def}}{=} (y_{i1\cdot\cdot}, y_{i2\cdot\cdot}, y_{i3\cdot\cdot}, y_{i4\cdot\cdot})'.$$

Then, collapsing over $\boldsymbol{\pi}$ and $\boldsymbol{\phi}$, we have that

$$h(\Omega,\boldsymbol{y}_{(1)}|\boldsymbol{y}_1,\boldsymbol{u},\boldsymbol{v},\boldsymbol{w}) \propto \frac{a_0}{(a_0+\tau_1)^2}\cdot\frac{b_0}{(b_0+\tau_2)^2}\cdot\frac{c_0}{(c_0+\beta)^2}$$
$$\times\prod_{i=1}^{A}\binom{n_i}{\boldsymbol{y}'_{i1},\,\boldsymbol{y}'_{i2},\,\boldsymbol{y}'_{i3},\boldsymbol{y}'_{i4}}\frac{\mathrm{D}(\boldsymbol{y}_i^{(1)}+\boldsymbol{\mu}_1\tau_1)}{\mathrm{D}(\boldsymbol{\mu}_1\tau_1)}\frac{\mathrm{D}(\boldsymbol{y}_i^{(2)}+\boldsymbol{\mu}_2\tau_2)}{\mathrm{D}(\boldsymbol{\mu}_2\tau_2)}$$
$$\times\prod_{i=1}^{A}\left[\frac{\beta^{\beta}}{\Gamma(\beta)}\right]^{rc}\mathrm{D}(\boldsymbol{y}_i^{(1)}+\beta\,\boldsymbol{j})\,I_i, \qquad (A.2)$$

where

$$I_i=\int\!\!\int\!\!\prod_{j,k}\left\{\frac{\left[\left(\frac{1-\pi_{i1}}{\phi_{ijk}}\right)\left(1+\frac{1}{\pi_{i1}}\{1-\pi_{i1}-\phi_{ijk}\}\right)\right]^{y_{i1jk}}}{(1-\pi_{i1})^{y_{i\cdot jk}+\beta}G_{\beta}\left(\frac{\beta}{1-\pi_{i1}}\right)}\right\}$$
$$\times\left\{\prod_{j,k}e^{-\frac{\beta\phi_{ijk}}{1-\pi_{i1}}}\right\}\left\{\frac{\prod_{j,k}\phi_{ijk}^{y_{i\cdot jk}+\beta-1}}{\mathrm{D}(\boldsymbol{y}_i^{(1)}+\beta\,\boldsymbol{j})}\right\}$$
$$\left\{\frac{\prod_{s=1}^{4}\pi_{is}^{y_{is\cdot\cdot}+\mu_{2s}\tau_2-1}}{\mathrm{D}(\boldsymbol{y}_i^{(2)}+\boldsymbol{\mu}_2\tau_2)}\right\}\mathrm{d}\boldsymbol{\phi}_i\,\mathrm{d}\boldsymbol{\pi}_i. \qquad (A.3)$$

Note that $0 \le \pi_{is} \le 1, \sum_{s=1}^{4} \pi_{is} = 1$ and $0 \le \phi_{ijk} \le 1$. We simplify the computation for I_i in (A.3) in two steps.

First, in (A.3) we make the transformation

$$\phi_{ijk} = T_i \, \phi^*_{ijk} \qquad \sum_{j=1}^{r} \sum_{k=1}^{c} \phi_{ijk} = T_i.$$

The new variables ϕ^*_{ijk} satisfy the relationships $0 \le \phi^*_{ijk} \le 1, \sum_{j=1}^{r} \sum_{k=1}^{c} \phi^*_{ijk} = 1$ and the T_i are restricted so that $0 \le T_i \le 1 / \phi^*_{ijk}$, for $j = 1, \ldots, r, k = 1, \ldots, c$ and $0 \le T_i \le rc$. With this transformation we have

$$I_i = \iiint_0^{b_i} \prod_{j,k} \left\{ \left[\frac{T_i}{1-\pi_{i1}} \right]^{\sum_{s=2}^{4} y_{isjk}} \left[\frac{1}{\phi^*_{ijk}} \left(1 + \frac{1-\pi_{i1}}{\pi_{i1}} \left\{ 1 - \frac{T_i}{1-\pi_{i1}} \phi^*_{ijk} \right\} \right) \right]^{y_{i1\,jk}} \right\}$$

$$\times \left\{ \frac{\left(\frac{T_i}{1-\pi_{i1}} \right)^{rc\beta - 1} e^{-\frac{\beta T_i}{1-\pi_{i1}}}}{(1-\pi_{i1}) \left[G_\beta \left(\frac{\beta}{1-\pi_{i1}} \right) \right]^{rc}} \right\} \left\{ \frac{\prod_{j,k} \phi^{*\, y_{i\cdot jk} + \beta - 1}_{ijk}}{\mathrm{D}(\boldsymbol{y}_i^{(1)} + \beta \boldsymbol{j})} \right\}$$

$$\times \left\{ \frac{\prod_{s=1}^{4} \pi_{is}^{y_{is\cdot} + \mu_{2s}\tau_2 - 1}}{\mathrm{D}(\boldsymbol{y}_i^{(2)} + \boldsymbol{\mu}_2 \tau_2)} \right\} \mathrm{d}T_i \, \mathrm{d}\boldsymbol{\phi}^*_i \mathrm{d}\boldsymbol{\pi}_i,$$

where $b_i = \min \{ \{1 / \phi^*_{ijk}\} \, j = 1, \ldots, r; k = 1, \ldots, c \}$.

Second, letting $W_i = \{\beta T_i / 1 - \pi_{i1}\}$ and absorbing the factor $\beta^{rc\beta} / \Gamma(rc\beta)$ in I_i, with some additional algebra, we have

$$I_i = \iiint_0^{\frac{\beta b_i}{1-\pi_{i1}}} \frac{G_{rc\beta} \left(\frac{\beta b_i}{1-\pi_{i1}} \right)}{\left[G_\beta \left(\frac{\beta}{1-\pi_{i1}} \right) \right]^{rc}}$$

$$\prod_{j,k} \left\{ \left(\frac{W_i}{\beta} \right)^{\sum_{s=2}^{4} y_{isjk}} \left[\frac{1}{\phi^*_{ijk}} \left(1 + \frac{1-\pi_{i1}}{\pi_{i1}} \left\{ 1 - \frac{W_i \phi^*_{ijk}}{\beta} \right\} \right) \right]^{y_{i1\,jk}} \right\}$$

$$\times \frac{W_i^{rc\beta - 1} e^{-W_i}}{\Gamma(rc\beta) G_{rc\beta} \left(\frac{\beta b_i}{1-\pi_{i1}} \right)}$$

$$\mathrm{d}W_i \left\{ \frac{\prod_{j,k} \phi^{*\, y_{i\cdot jk} + \beta - 1}_{ijk}}{\mathrm{D}(\boldsymbol{y}_i^{(1)} + \beta \boldsymbol{j})} \right\} \left\{ \frac{\prod_{s=1}^{4} \pi_{is}^{y_{is\cdot} + \mu_2 s \tau_2 - 1}}{\mathrm{D}(\boldsymbol{y}_i^{(2)} + \boldsymbol{\mu}_2 \tau_2)} \right\} \mathrm{d}\boldsymbol{\phi}^*_i \, \mathrm{d}\boldsymbol{\pi}_i. \quad \text{(A.4)}$$

Appendix B

Ignorable and nonignorable nonresponse models

Set $\psi_{ijk} \equiv 1$ in the expansion model to form the ignorable nonresponse model. For $i = 1, \ldots, A,$ we then take

$$\boldsymbol{\pi}_i \,|\, \boldsymbol{\mu}_2, \tau_2 \overset{\text{iid}}{\sim} \text{Dirichlet}(\boldsymbol{\mu}_2 \tau_2)$$

and independently

$$\boldsymbol{p}_i \,|\, \boldsymbol{\mu}_1, \tau_1 \overset{\text{iid}}{\sim} \text{Dirichlet}(\boldsymbol{\mu}_1 \tau_1).$$

Also, $p(\tau_2) = \{1/(1+\tau_1)^2\}, \tau_1 \ge 0, \mu_1 \sim \text{Dirichlet}(\mathbf{1})$, $p(\tau_2) = \{1/(1+\tau_2)^2\}, \tau_1 \ge 0$ and $\mu_2 \sim \text{Dirichlet}(\mathbf{1})$. Here we have independence at all levels and the vectors $\mathbf{1}$ are of the appropriate dimension with every coordinate equal to one. Note, that all the parameters of the ignorable model are identifiable and estimable.

Set $\pi_{isjk} = \pi_{is} \psi_{ijk}$ in the expansion model to form the nonignorable nonresponse model. In this case, for $i = 1, \ldots, A,$

$$\boldsymbol{\pi}_{ijk} \,|\, \boldsymbol{\mu}_2, \tau_2 \overset{\text{iid}}{\sim} \text{Dirichlet}(\boldsymbol{\mu}_2 \tau_2)$$

and independently

$$\boldsymbol{p}_i \,|\, \boldsymbol{\mu}_1, \tau_1 \overset{\text{iid}}{\sim} \text{Dirichlet}(\boldsymbol{\mu}_1 \tau_1).$$

In this model, the parameters $\boldsymbol{\pi}_{ijk}$ are not identifiable and we take $\tau_2 \sim \text{Gamma}(\alpha_0, \beta_0)$, where α_0 and β_0 are to be specified. The model specification is then completed by assigning $\tau_1, \boldsymbol{\mu}_1$ and $\boldsymbol{\mu}_2$ the same distributional properties as in the previous paragraph.

As in Nandram, Cox and Choi (2005), α_0 and β_0 are specified as follows. The ignorable nonresponse model is fit to obtain a sample from the posterior density of τ_2. Then α_0 and β_0 are obtained using the method of moments. Nandram, Cox and Choi (2005) found that inference about $\boldsymbol{p}_i$ is not very sensitive to the choice of these parameters.

References

Cohen, G., and Duffy, J.C. (2002). Are nonrespondents to health surveys less healthy than respondents? *Journal of Official Statistics*, 18, 13-23.

Draper, D. (1995). Assessment and propagation of model uncertainty (with discussion). *Journal of the Royal Statistical Society*, Series B, 57, 45-97.

Forster, J.J., and Smith, P.W.F. (1998). Model-based inference for categorical survey data subject to non-ignorable nonresponse. *Journal of the Royal Statistical Society*, Series B, 60, 57-70.

Gelman, A. (2006). Prior distribution for variance parameters in hierarchical models. *Bayesian Analysis*, 1, 515-533.

Greenland, S. (2009). Relaxation penalties and priors for plausible modeling of nonidentified bias sources. *Statistical Sciences*, 24, 195-210.

Kass, R.E., and Raftery, A.E. (1995). Bayes factors. *Journal of the American Statistical Association*, 90, 773-795.

Little, R.J.A., and Rubin, D.B. (2002). *Statistical Analysis with Missing Data*, Second Edition. New York: John Wiley & Sons, Inc.

Nandram, B. (2009). Bayesian inference of the cell probabilities of a two-way categorical table under non-ignorability. *Communications in Statistics - Theory and Methods*, 38, 3015-3030.

Nandram, B., and Choi, J.W. (2002a). Hierarchical Bayesian nonresponse models for binary data from small areas with uncertainty about ignorability. *Journal of the American Statistical Association*, 97, 381-388.

Nandram, B., and Choi, J.W. (2002b). A Bayesian analysis of a proportion under nonignorable nonresponse. *Statistics in Medicine*, 21, 1189-1212.

Nandram, B., and Choi, J.W. (2004). A nonparametric Bayesian analysis of a proportion for a small area under nonignorable nonresponse. *Journal of Nonparametric Statistics*, 16, 821-839.

Nandram, B., and Choi, J.W. (2005). Hierarchical Bayesian nonignorable nonresponse regression models for small areas: An application to the NHANES data. *Survey Methodology*, 31, 73-84.

Nandram, B., and Choi, J.W. (2008). A Bayesian allocation of undecided voters. *Survey Methodology*, 34, 37-49.

Nandram, B., and Choi, J.W. (2010). A Bayesian analysis of body mass index data from small domains under nonignorable nonresponse and selection. *Journal of the American Statistical Association*, 105, 120-135.

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

Nandram, B., Han, G. and Choi, J.W. (2002). A hierarchical Bayesian nonignorable nonresponse model for multinomial data from small areas. *Survey Methodology*, 28, 145-156.

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. *Statistics in Medicine*, 24, 1047-1074.

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

Rubin, D.B. (1976). Inference and missing data. *Biometrika*, 63, 581-592.

Smith, A.F.M., and Gelfand, A.E. (1992). Bayesian statistics without tears: A sampling-resampling perspective. *The American Statistician*, 46, 84-88.