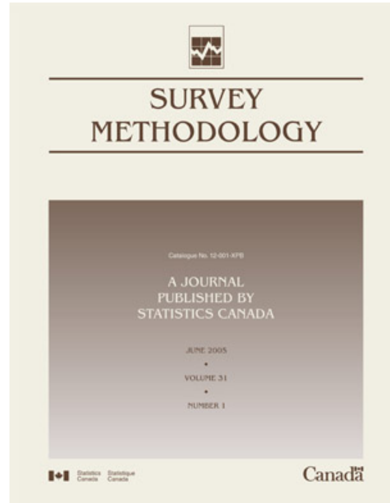
Catalogue no. 12-001-XIE

Survey Methodology

December 2002



Statistics Canada Statistique Canada

Canada

How to obtain more information

Specific inquiries about this product and related statistics or services should be directed to: Business Survey Methods Division, Statistics Canada, Ottawa, Ontario, K1A 0T6 (telephone: 1 800 263-1136).

For information on the wide range of data available from Statistics Canada, you can contact us by calling one of our toll-free numbers. You can also contact us by e-mail or by visiting our website.

National inquiries line	1 800 263-1136
National telecommunications device for the hearing impaired	1 800 363-7629
Depository Services Program inquiries	1 800 700-1033
Fax line for Depository Services Program	1 800 889-9734
E-mail inquiries	infostats@statcan.ca
Website	www.statcan.ca

Information to access the product

This product, catalogue no. 12-001-XIE, is available for free. To obtain a single issue, visit our website at www.statcan.ca and select Our Products and Services.

Standards of service to the public

Statistics Canada is committed to serving its clients in a prompt, reliable and courteous manner and in the official language of their choice. To this end, the Agency has developed standards of service that its employees observe in serving its clients. To obtain a copy of these service standards, please contact Statistics Canada toll free at 1 800 263-1136. The service standards are also published on www.statcan.ca under About Statistics Canada > Providing services to Canadians.

Statistics Canada
Business Survey Methods Division

Survey Methodology

December 2002

Published by authority of the Minister responsible for Statistics Canada

January 2006

Catalogue no. 12-001-XIE
ISSN 1492-0921

Frequency: semi-annual

Ottawa

Cette publication est disponible en français sur demande (nº 12-001-XIF au catalogue).

Note of appreciation

Canada owes the success of its statistical system to a long-standing partnership between Statistics Canada, the citizens of Canada, its businesses, governments and other institutions. Accurate and timely statistical information could not be produced without their continued cooperation and goodwill.

Vol. 28, No. 2, pp. 145-156
Statistics Canada, Catalogue No. 12-001

A Hierarchical Bayesian Nonignorable Nonresponse Model for Multinomial Data from Small Areas

Balgobin Nandram, Geunshik Han and Jai Won Choi [1]

Abstract

The analysis of survey data from different geographical areas, where the data from each area are polychotomous, can be easily performed using hierarchical Bayesian models even if there are small cell counts in some of these areas. However, there are difficulties when the survey data have missing information in the form of nonresponse especially when the characteristics of the respondents differ from the nonrespondents. We use the selection approach for estimation when there are nonrespondents because it permits inference for all the parameters. Specifically, we describe a hierarchical Bayesian model to analyze multinomial nonignorable nonresponse data from different geographical areas, some of them can be small. For the model, we use a Dirichlet prior density for the multinomial probabilities and a beta prior density for the response probabilities. This permits 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. Because the joint posterior density of all the parameters is complex, inference is sampling based and Markov chain Monte Carlo methods are used. We apply our method to provide an analysis of body mass index (BMI) data from the third National Health and Nutrition Examination Survey (NHANES III). For simplicity, the BMI is categorized into three natural levels, and this is done for each of eight age-race-sex domains and thirty-four counties. We assess the performance of our model using the NHANES III data and simulated examples, which show our model works reasonably well.

Key Words: Latent variable; Metropolis-Hastings sampler; Nonignorable nonresponse; Selection approach; Small area.

1. Introduction

The nonresponse rates in many surveys have been increasing steadily (De Heer 1999; Groves and Couper 1998), making the nonresponse problem more important For many surveys the responses are polychotomous. For example, in the third National Health and Nutrition Examination Survey (NHANES III), we can estimate the proportions of persons belonging to three levels of body mass index (BMI), although BMI is a continuous variable. The purpose of this paper is to describe a new hierarchical Bayesian model to study nonignorable multinomial nonresponse for small areas, and to apply it to the NHANES III BMI data.

Rubin (1987) and Little and Rubin (1987) describe two types of models which differ according to the ignorability of response. In the ignorable nonresponse model the distribution of the variable of interest for a respondent is the same as the distribution of that variable for a nonrespondent with the same values of the covariates. In addition, the parameters in the distributions of the variable and response must be distinct (see Rubin 1976). All other nonresponse models are nonignorable. We use both ignorable and nonignorable nonresponse models for our data because there are no nonrespondents for some domains.

Crawford, Johnson and Laird (1993) used nonignorable nonresponse models to analyze data from the Harvard Medical Practice Survey. Stasny, Kadane, and Fritsch (1998) used a Bayesian hierarchical model for the probabilities of voting guilty or not on a particular trial when the views of nonrespondents differ from those of respondents in various death-penalty beliefs. Park and Brown (1994) used a pseudo-Bayesian method (Baker and Laird 1988), and Park (1998) applied a method in which prior observations are assigned to both observed and unobserved cells to estimate the missing cells of a multi-way categorical table under nonignorable nonresponse. Our approach differs from these authors. We describe small area estimation for multinomial data, and we use Markov chain Monte Carlo methods to implement the methodology. This permits the inclusion of all sources of variability in our models.

There are two approaches to model nonresponse. The selection approach is used for the hypothetical complete data, and a nonresponse model is added conditional on the hypothetical data. This approach was developed to study sample selection problems (*e.g.*, Heckman 1976 and Olson 1980). In the pattern mixture approach the respondents and the nonrespondents are modeled separately, and the final answer is obtained by a probabilistic mixture of the two. We use the selection approach for our problem.

Stasny (1991) used an empirical Bayes model to study victimization in the National Crime Survey, and she followed the selection approach. This analysis pools binomial data from several domains, and some of them have small counts. Essentially this is an exercise in small area estimation. A related method was presented by Albert and

1. Balgobin Nandram, Department of Mathematical Sciences, Worcester Polytechnic Institute, 100 Institute Road, Worcester, MA 01609-2280, U.S.A. E-mail: balnan@wpi.edu; Geunshik Han, Division of Computer and Information Science, Hanshin University, Osan, Korea. E-mail: gshan@hucc.hanshin.ac.kr; Jai Won Choi, National Center for Health Statistics, Room 915, 6525 Belcrest Road, Hyattsville, MD 20782, U.S.A. E-mail: jwc7@cdc.gov.

Gupta (1985), who used an approximation to obtain a Bayesian approach for a population with a single domain (see also Kaufman and King 1973). That is, unlike Stasny (1991), these latter authors did not perform. small area estimation, and their analysis in a single domain do not use data from other domains.

Since the Bayesian approach can incorporate other information about nonrespondents, the Bayesian method is appropriate for the analysis of nonignorable nonresponse (Little and Rubin 1987 and Rubin 1987). However the main difficulty is how to describe the relationship between the respondents and nonrespondents. Using the selection approach within the framework of Bayes empirical Bayes (see Deely and Lindley 1981), Stasny (1991) estimated the hyper-parameters by maximum likelihood methods and then assumed them known, thereby suppressing some variability. We extend this approach in two directions.

First, we consider multinomial data obtained independently from several geographical areas. It is worthy to note that Basu and Pereira (1982) considered multinomial nonresponse data from a single domain using a multinomial Dirichlet model when the hyper-parameters are assumed known. Recently, Forster and Smith (1998) used graphical multinomial Dirichlet log-linear models to analyze data from the panel survey in British general election. Again the hyper-parameters are assumed known, and a model with a single domain is used. Secondly, we obtain a full Bayesian approach for multinomial nonignorable nonresponse data from several areas. We do not estimate the hyper-parameters using the data.

As a summary, we develop a multinomial nonignorable nonresponse model which is used for pooling data over many small areas, and we note that it can be used in other applications. The rest of the paper is organized as follows. In section 2 we describe the NHANES III. In section 3 we discuss the Bayesian model for nonignorable nonresponse. In particular, a three-stage Bayesian hierarchical multinomial model is applied to the NHANES III data to investigate the nonresponse problem. In section 4 we describe an analysis of the NHANES III data in which we include a regression analysis to combine all the age-race-sex domains. In section 5 we describe a simulation study to assess the performance of our model. Finally, section 6 has the conclusion.

2. NHANES III Data and Nonresponse

The NHANES III is one of the periodic surveys used to assess an aspect of health of the U.S. population (National Center for Health Statistics 1994). Our research is motivated by nonresponse of body mass index (BMI) in the NHANES III. The data for our illustration come from this survey, and were collected from October 1988 to September 1994. In section 2.1 we describe the actual data, and in section 2.2 we describe the data we analyze.

2.1 NHANES III Data

The NHANES III consists of two parts. The first part is the interview of the sampled individuals for their personal information and the second part is the examination of those sampled. One or more persons from the sampled households were placed into a number of subgroups depending on their age, race and sex. Some subgroups were sampled at different rates. Sampled persons were asked to come to a mobile examination center (MEC) for a phyzsical examination, Those who did not come were visited by the examiner for the same purpose. Details of the NHANES III sample design are available (National Center for Health Statistics 1992). We incorporate design features associated with clustering in our model.

The main reasons for NHANES III nonresponse are “not interested”, “no time/work conflict”, “concerns/suspicious”, “don’t bother me” and “health reasons”. The nonresponse rate of younger individuals is very high because the parents, especially older mothers of an only child, were extremely protective of their babies, and would not allow them to leave their homes for the MECs. Field workers often observe that obese persons tend to avoid the medical examination. So that nonresponse might be nonrandom and hence require some special attention.

NHANES III data are adjusted by multistage ratio weightings for the data to be consistent with the population (Mohadjer, Bell and Waksberg 1994). The ratio is the proportion of persons in the sample to the number of persons who completed interview and examination. Weighting with nonresponse ratio is one of these stages. In nonresponse ratio estimation, the proportions of nonrespondents in the multinomial cells are the same as those for the respondents (*i.e.*, ignorable nonresponse). In this case since the proportions are of interest, no adjustment is required. Clearly, this ratio estimation can be incorrect when these two groups are different. Therefore there is a need to consider the adjustment by a method other than ratio adjustment. In this paper we investigate a Bayesian method as an alternative to ratio weighting for nonignorable nomesponse.

NHANES III nonresponse also occurs at several levels in the survey: interview and examination. The interview nonresponse arises from sample individuals who did not respond for the interview. Some of those who were already interviewed did not come to the MEC, missing all or part of the examinations. In this paper, our population consists of those individuals who would have agreed to take the physical examination in the MECs. Thus, nonrespondents are those individuals who agreed to take the physical examination, and did not show up at the MECs. More specifically, since we are considering item response, the nonrespondents are those individuals who agreed to come to the MECs and their heights and/or weights were not measured.

Schafer, Ezzati-Rice, Johnson, Khare, Little and Rubin (1996) attempted a comprehensive multiple imputation project on the NHANES III data for many variables. The purpose was to impute the nonresponse data to provide several data sets for public use. Unfortunately, one of the limitations of the project was that "the procedure used to create missingness corresponds to a purely ignorable mechanism; the simulation provides no information on the impact of possible deviations from ignorable nonresponse." Another limitation is that the procedure did not include geographical clustering. Our purpose is different; we do not provide imputed public-use data.

2.2 Data Used for Illustration

Our data have two age groups (younger than 45 years, 45–, and 45 years or older, 45+), two race groups (white and non-white) and, of course, two groups for sex (male and female). Thus, there are eight age-race-sex domains.

One of the variables of interest in the NHANES III is BMI, an index of weight adjusted for height (Kg/m^2) that broadly categorizes obesity within age-race-sex groups (Kuczmarski, Carrol, Flegal and Troiano 1997) as low body fat (level 1: BMI < 20), healthy body fat (level 2: 20 ≤ BMI < 25), hefty or unhealthy (level 3: BMI ≥ 25). We use this broad classification for each of the eight age-race-sex groups.

Rather than a categorical data analysis, one can also provide an analysis that treats BMI as a continuous variable. While some information is lost by discretizing the BMI values, an analysis using continuous models for BMI will also be approximate and there is a need to search for an appropriate transformation. In the final analysis, a doctor only needs to know what proportions of the public belong to different levels of BMI, so he or she can tell his patient's standing in obesity.

The analysis of BMl data using categorical data methods is not uncommon. For example, Malec, Davis and Cao (1999) described a Bayes empirical Bayes analysis of the NHANES III data. They classified an individual older than 20 years as normal if her/his BMI is below a certain gender specific threshold. This is an application of a Bayesian analysis of binary data. However, their classification is somewhat restricted (see Kuczmarski *et al.* 1997). By considering multinomial data, we have generalized the analysis of Malec *et al.* (1999). In fact, they did not provide a nonignorable nonresponse model.

Unlike Schafer *et al.* (1996), we include clustering at the county level, although there is a need to include clustering at the household level. For the complete data there are 6,440 households. Of these households 52.1% contributed one person to the sample, 22.5% two persons, and 21.4% at least three persons. We have calculated the correlation coefficient for the BMI values based on pairing the members within households (see Rao 1973, page 199). It is 0.19 which indicates that as a first approximation the clustering within households can be ignored.

Table 1 shows the number of respondents for each BMI level for each age-race-sex domain and 34 counties (population at least 500,000). The pattern of respondents differs greatly by age. The nonresponse rate for the older group (45+) is negligible. Therefore the main concern about nonresponse must be given to the younger group (45–). There is also higher response rate among females than males. We note that the selection procedure is not random over the single population of males and females.

Table 1
Number of Individuals in Each BMI Level and Number of Nonrespondents (Non) by Age, Race and Sex Over All 34 Counties

Age	Race	Sex	BMI 1	BMI 2	BMI 3	Non
45–	W	M	1,098	651	597	558
		F	845	434	380	233
	B	M	1,198	713	665	574
		F	745	463	524	214
45+	W	M	46	439	1,014	3
		F	51	223	365	4
	B	M	79	470	942	8
		F	48	169	552	6

Note: BMI (1=less than 20; 2 = at least 20 and smaller than 25;
3 = greater than 25)
Age (Younger than 45 years = 45–; 45 years or older = 45+)
Race (White = W; all others = B)
Sex (Male = M; Female = F).

Table 2
Number of Individuals in Each BMI Level and Number of Nonrespondents (Non) for Eight Examples (Ex) of Small Age-race-sex Domains from Different Counties

Ex	Age	Race	Sex	BMI Level 1	BMI Level 2	BMI Level 3	Non
1	45–	W	M	1	3	1	14
2			F	3	4	1	0
3		B	M	5	5	6	10
4			F	3	1	1	1
5	45+	W	M	1	2	6	0
6			F	1	3	4	0
7		B	M	3	3	5	0
8			F	2	0	1	1

Note: BMI (1=less than 20; 2 = at least 20 and smaller than 25;
3 = greater than 25)
Age (Younger than 45 years = 45–; 45 years or older = 45+)
Race (White = W; all others = B)
Sex (Male = M; Female = F).

One important aspect of our work is on small area estimation. Because we consider inference for each age-race-sex domain separately over the the geographical areas (counties), the samples from some of these areas can be very small. Thus, small area estimation techniques are required to estimate the parameters corresponding to these smaller areas. Specifically, we need to "borrow strength" from the larger areas to make the estimates for the smaller areas more reliable. Table 2 presents eight examples to show the need for small area techniques. We have selected eight counties that have small domains; all the cell counts are at most 6 and many of them are as small as 1 (one of them is 0 for 45+). We will present overall estimates and the estimates for the first four examples (45–). Note that in comparison to the cell counts, the nonrespondents are large for two of them (14 and 10 nonrespondents).

We note that the purpose is not a comprehensive analysis of the NHANES III data although it forms an approximate analysis for these data. Our method is general enough to analyze multinomial nonresponse data from many areas, some of which can be small. It is for these small areas that we develop this modeling technique. Thus, in this paper we use the NHANES III data to illustrate our method.

Our method considers each domain separately with a "borrowing of strength" across the 34 areas (counties) to analyze the BMI data. Thus, there are eight separate analyses, each with 34 areas, and some of them are small. We use a hierarchical multinomial nonresponse model to analyze data of this form. The small cell counts, substantial nonrespondents and multinomial data make the methodology much more practical. Our methodology is also extended to incorporate all the domains simultaneously through logistic models.

3. Methodology for Hierarchical Multinomial Model

We propose a model for each of the eight age-race-sex domains but for all counties taken simultaneously. However, the models fall into two broad classes. We will use a nonignorable nonresponse model for the younger group and an ignorable nonresponse model for the older group since the nonresponse rate for the older group is negligible. Of course, it is worthwhile to compare the ignorable nonresponse model and the nonignorable nonresponse model for the younger group. We will show how to combine the groups later using logistic regression, although this is not the key issue of this paper.

For each age-race-sex group, the k^{th} individual in the i^{th} county belongs to one of J BMI levels. Then for the k^{th} individual in i^{th} county, the characteristic variable at the j^{th} BMI level is defined as follows,

$$\mathrm{x}_{ik} = \left(x_{i1k}, \ldots, x_{ijk}, \ldots, x_{iJk}\right)', \; i = 1, \ldots, c, \; k = 1, \ldots, n_i,$$

where each $x_{ijk} = 0$ or 1, $j = 1, \ldots, J$, and $\sum_{j=1}^{J} x_{ijk} = 1$. The response variable, y_{ijk}, is defined for each age-race-sex domain

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

We use a probabilistic structure to model the $\mathbf{x}_{ik}$ and y_{ijk}. In our application, there are c = 34 counties and J = 3 BMI levels.

3.1 Ignorable and Nonignorable Nonresponse Models

For both ignorable and the nonignorable nonresponse models, we have

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

where p_{ij} is the probability that an individual in the i^{th} county belongs the j^{th} BMI level. Next, we describe the remaining portions of the ignorable and the nonignorable models.

First, we describe the ignorable nonresponse model. Let π_i denote the probability that an individual within the i^{th} county responds (*i.e.*, the probability of responding depends only on the county). Then, we assume that

$$y_{ijk} \mid \boldsymbol{\pi}_i \overset{\text{iid}}{\sim} \text{Bernoulli}\,(\boldsymbol{\pi}_i). \qquad (2)$$

At the second stage, letting $\boldsymbol{\mu}_1 = (\mu_{11}, \mu_{12}, \ldots, \mu_{1J})'$, we take

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

$$\boldsymbol{\pi}_i \mid \mu_{21},\ \tau_{21} \overset{\text{iid}}{\sim} \text{Beta}\,(\mu_{21}\ \tau_{21}, (1-\mu_{21})\tau_{21}) \qquad (4)$$

where

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

and

$$D(\boldsymbol{\mu}_1\tau_1) = \prod_{j=1}^{J} \Gamma(\mu_{1j}\tau_1)/\Gamma(\tau_1), 0 < \mu_{1j} < 1, \sum_{j=1}^{J} \mu_{1j} = 1.$$

The components of $\boldsymbol{\mu}_1$ are the prior means of the corresponding components of the $\mathbf{p}_i$, and τ_1 can be interpreted as a prior sample size. Similar interpretations can be given for μ_{21} and τ_{21} for $\boldsymbol{\pi}_i$. Thus, assumption (3) expresses similarity among the cell proportions $\mathbf{p}_i$ and (4) expresses similarity among the response probabilities $\boldsymbol{\pi}_i$. It is this structure that causes the "borrowing of strength" across the c counties.

Second, we describe the nonignorable nonresponse model. Let $\boldsymbol{\pi}_{ij}$ denote the probability that an individual within the i^{th} county responds in the j^{th} BMI level (*i.e.*, the probability of responding depends not only on the county but also on the BMI level). Then, we assume that

$$y_{ijk} \mid \{\mathrm{x}_{ik} = (x_{i1k}, \ldots, x_{iJk}), \boldsymbol{\pi}_{ij}\} \overset{\text{iid}}{\sim} \text{Bernoulli}\,(\boldsymbol{\pi}_{ij}) \quad (5)$$

where $x_{ijk} = 1, x_{ij'k} = 0, j \neq j'$ for $j, j' = 1, 2, \ldots, J$. Letting $\boldsymbol{\mu}_3 = (\mu_{31}, \mu_{32}, \ldots, \mu_{3J})'$, at the second stage we also take

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

and

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

Like the assumptions in (3) and (4), the assumptions in (6) and (7) express similarity among the counties. We note that the response parameters π_{ij} are weakly identifiable (*i.e.*, unreliable estimates). However, the selection model works to our advantage, because the joint density of $\mathbf{x}_{ik}$ and $\mathbf{y}_{ik} = (y_{i1k}, \ldots, y_{iJk})'$ connects the p_{ij} and π_{ij}. In fact, this is an advantage over the pattern mixture approach.

To ensure a full Bayesian analysis, at the third stage we take the prior densities for the hyper-parameters as follows. For the ignorable nonresponse model, the prior densities are

$$\boldsymbol{\mu}_1 \sim \text{Dirichlet}\,(1,1,\ldots,1),\ \mu_{21} \sim \text{Beta}\,(1,1),$$
$$\tau_1 \sim \text{Gamma}\,(\eta_1^{(0)}, v_1^{(0)}) \text{ and}$$
$$\tau_{21} \sim \text{Gamma}\,(\eta_{21}^{(0)}, v_{21}^{(0)}),$$

where (letting t denote either τ_1 or τ_{21}, a either $\eta_1^{(0)}$ or $\eta_{21}^{(0)}$, and b either $v_1^{(0)}$ or $v_{21}^{(0)}$ $\tau \sim \text{Gamma}\,(a,b)$ means that $f(t) = b^a t^{a-1} e^{-bt} / \Gamma(a),\ t > 0$ and $f(t) = 0$ otherwise. The hyper-parameters $\eta_1^{(0)}, v_1^{(0)}, \eta_{21}^{(0)}$ and $v_{21}^{(0)}$ are to be specified. The corresponding part of the nonignorable nonresponse model is

$$\boldsymbol{\mu}_3 \sim \text{Dirichlet}\,(1,1,\ldots,1),\ \mu_{4j} \sim \text{Beta}\,(1,1),$$
$$\tau_3 \sim \text{Gamma}\,(\eta_3^{(0)}, v_3^{(0)}) \text{ and}$$
$$\tau_{4j} \sim \text{Gamma}\,(\eta_{4j}^{(0)}, v_{4j}^{(0)}),\ j = 1, \ldots, J.$$

Again, the hyper-parameters $\eta_3^{(0)}, v_3^{(0)}, \eta_{4j}^{(0)}, v_{4j}^{(0)}$, $j = 1, \ldots, J$, are to specified. It is possible to use other prior densities such as shrinkage priors, but it is likely that these will provide similar inference as our sensitivity analysis indicates in section 4.

It is an attractive property of the hierarchical model that it introduces correlation among the variables. For example, in our application (1), (2), (3) and (4) make the $(\mathbf{x}_{ij}, y_{ij})$ equi-correlated across the individuals within the i^{th} area. This is the clustering effect within the areas. Such an effect can be obtained directly, but it will not be as simple as in a hierarchical model. A further benefit of the hierarchical model is that it takes care of extraneous variations among the areas, and this effect can be obtained directly by using random effects model. But in our case, this will loose the natural multinomial data structure.

Let r_i be the number of respondents in county i and y_{ij} the number of respondents having the j^{th} BMI level in the i^{th} county. Then r_i and y_{ij} are random variables; $n_i - r_i$ is the number of nonrespondents. Since the number of nonrespondents at the j^{th} BMI level is unknown, we denote them by the latent variables z_{ij} (see the tree diagram in Figure 1). If we can tell what the z_{ij} are, our nonresponse problem will be solved. Of course, under the assumption of ignorable nonresponse, they can be estimated easily using ratio estimation. The z_{ij} are useful because under the assumption of nonignorable nonresponse they simplify the sampling based method to obtain estimates of the parameters of interest.

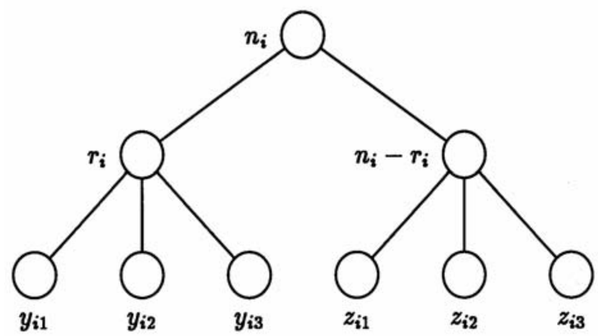


Figure 1. Latent nonignorable response tree diagram. From a sample of n_i individuals, there are r_i respondents of which y_{ij} belong to category j, $j = 1, 2, 3$. Among the $(n_i - r_i)$ nonrespondents z_{ij} individuals belong to category j, where z_{ij} are latent variables.

The likelihood function for the ignorable nonresponse model is

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

Here the likelihood function has two distinct parts, one for p_{ij} and the other for the π_i. Using Bayes' theorem the joint posterior density of all the parameters is

$$f(\mathbf{p}, \boldsymbol{\pi}, \boldsymbol{\mu}_1, \tau_1, \mu_{21}, \tau_{21} \mid \mathbf{y}, \mathbf{r})$$
$$\propto \prod_{i=1}^{c} \left\{ \left\{ \prod_{j=1}^{J} p_{ij}^{\,y_{ij} + n_i - r_i} \right\} \left\{ \pi_i^{r_i} (1 - \pi_i)^{n_i - r_i} \right\} \right.$$
$$\times \left\{ \prod_{j=1}^{J} p_{ij}^{\,\mu_{1j}\tau_1 - 1} \right\} \Big/ D(\boldsymbol{\mu}_1 \tau_1)$$
$$\left. \times \left\{ \frac{\pi_i^{\mu_{21}\tau_{21} - 1} (1 - \pi_i)^{(1-\mu_{21})\tau_{21} - 1}}{B(\mu_{21}\tau_{21}, (1 - \mu_{21})\tau_{21})} \right\} \right\}$$
$$\times \left\{ \tau_1^{\eta_1^{(0)} - 1} \exp(-v_1^{(0)} \tau_1) \right\} \left\{ \tau_{21}^{\eta_{21}^{(0)} - 1} \exp(-v_{21}^{(0)} \tau_{21}) \right\}. \quad (8)$$

Similarly, the augmented likelihood function (*i.e.*, including the $\mathbf{z}_i$) for the nonignorable nonresponse model is

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

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

$$f(\mathbf{p}, \boldsymbol{\pi}, \mathbf{z}, \boldsymbol{\mu}_3, \tau_3\ \boldsymbol{\mu}_4, \boldsymbol{\tau}_4 \mid \mathbf{y}, \mathbf{r})$$

$$\propto \prod_{i=1}^{c} \left\{ \binom{n_i - r_i}{z_{i1}, .., z_{iJ}} \prod_{j=1}^{J} (\pi_{ij}\ p_{ij})^{y_{ij}} ((1-\pi_{ij})\ p_{ij})^{z_{ij}} \right.$$

$$\times \prod_{j=1}^{J} p_{ij}^{\mu_{3j}\tau_3 - 1} \Big/ D(\boldsymbol{\mu}_3\ \tau_3) \prod_{j=1}^{J}$$

$$\left. \times \left\{ \frac{\pi_{ij}^{\mu_{4j}\pi_{4j}-1} (1-\pi_{ij})^{(1-\mu_{4j})\tau_{4j}-1}}{B(\mu_{4j}\ \tau_{4j}, (1-\mu_{4j})\ \tau_{4j})} \right\} \right\}$$

$$\times \left\{ \tau_3^{\eta_2^{(0)}-1} \exp(-v_2^{(0)}\ \tau_3) \right\} \prod_{j=1}^{J} \left\{ \tau_{4j}^{\eta_{4j}^{(0)}-1} \exp(-v_{4j}^{(0)}\ \tau_{4j}) \right\}.$$

We consider inference about the p_{ij}, the proportion of individuals at the j^{th} BMI level in the i^{th} county, and the probability of responding,

$$\delta_i = \sum_{j=1}^{J} \pi_{ij}\ p_{ij},\ i = 1, ..., c.$$

However, the joint posterior densities in (8) and (9) are complex, and can not be used to make inference analytically. Thus, we use a Markov chain Monte Carlo algorithm to obtain estimates of the posterior distribution of the parameters. Our method is to use a Metropolis-Hastings (MH) sampler to get samples from (8) and (9) and then to use these samples to make posterior inferences about $\mathbf{p}_i$ and δ_i.

3.2 Computations

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

$$f(\mathbf{p}, \boldsymbol{\pi}, \boldsymbol{\mu}_1, \tau_1, \mu_{21}, \tau_{21} \mid \mathbf{y}, \mathbf{r})$$

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

$$\times f_3(\boldsymbol{\mu}_1, \tau_1, \boldsymbol{\mu}_{21}, \tau_{21} \mid \mathbf{y}, \mathbf{r})$$

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

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

$f_2(\cdot)$ is beta density,

$$\boldsymbol{\pi}_i | \mathbf{y}_i, \mathbf{r}_i, \mu_{21}, \tau_{21} \overset{\text{ind}}{\sim}$$

$$\text{Beta}(r_i + \mu_{21}\ \tau_{21}, n_i - \boldsymbol{r}_i + (1-\mu_{21})\ \tau_{21})$$

and

$$f_3(\boldsymbol{\mu}_1, \tau_1, \boldsymbol{\mu}_{21}, \tau_{21} \mid \mathbf{y}, \mathbf{r})$$

$$\propto \prod_{i=1}^{c} \{ D(\mathbf{y}_i + n_i - r_i + \boldsymbol{\mu}_1\ \tau_1) / D(\boldsymbol{\mu}_1\ \tau_1) \}\ p(\boldsymbol{\mu}_1, \tau_1)$$

$$\times \prod_{i=1}^{c} \left\{ \frac{B(r_i + \mu_{21}\ \tau_{21}, n_i - r_i + (1-\mu_{21})\tau_{21})}{B(\mu_{21}\ \tau_{21}, (1-\mu_{21})\ \tau_{21})} \right\} p(\mu_{21}, \tau_{21})$$

with $p(\boldsymbol{\mu}_1, \tau_1)$ and $p(\mu_{21}, \tau_{21})$ the prior distributions. Hence, f_1 and f_2 are obtained through the Gibbs kernel, while for f_3 we use the MH algorithm (Nandram 1998).

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

$$f(\mathbf{p}, \boldsymbol{\pi}, \mathbf{z}, \boldsymbol{\mu}_3, \tau_3, \boldsymbol{\mu}_4, \boldsymbol{\tau}_4 \mid \mathbf{y}, \mathbf{r})$$

$$= \prod_{i=1}^{c} \left\{ \left\{ \prod_{j=1}^{J} f_j(\boldsymbol{\pi}_{ij} \mid \mathbf{y}, \mathbf{r}, \mathbf{z}, \mu_{4j}, \tau_{4j}) \right\} f_{J+1}(\mathbf{p}_i \mid \mathbf{y}, \mathbf{r}, \mathbf{z}, \boldsymbol{\mu}_3, \tau_3) \right\}$$

$$\times f_{J+2}(\boldsymbol{\mu}_3, \tau_3, \boldsymbol{\mu}_4, \boldsymbol{\tau}_4, \mathbf{z} \mid \mathbf{y}, \mathbf{r}),$$

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

$$\pi_{ij} \mid y_{ij}, r_{ij}, z_{ij}, \mu_{4j}, \tau_{4j} \overset{\text{ind}}{\sim}$$

$$\text{Beta}(y_{ij} + \mu_{4j}\ \tau_{4j}, z_{ij} + (1-\mu_{4j})\ \tau_{4j}),$$

$f_{J+1}(.)$ is a Dirichlet density,

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

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

$$f_{J+2}(\boldsymbol{\mu}_3, \tau_3, \mu_4, \tau_4, \mathbf{z} \mid \mathbf{y}, \mathbf{r})$$

$$\propto \prod_{i=1}^{c} \left\{ \binom{n_i - r_i}{z_{i1}, ..., z_{iJ}} \{ D(\mathbf{y}_i + \mathbf{z}_i + \boldsymbol{\mu}_3\ \tau_3) / D(\boldsymbol{\mu}_3\ \tau_3) \} \right.$$

$$p(\boldsymbol{\mu}_3, \tau_3)$$

$$\left. \times \prod_{j=1}^{J} \left\{ \frac{B(y_{ij} + \mu_{4j}\tau_{4j}, z_{ij} + (1-\mu_{4j})\tau_{4j})}{B(\mu_{4j}\ \tau_{4j}, (1-\mu_{4j})\tau_{4j})} \right\} \right\} p(\boldsymbol{\mu}_4, \boldsymbol{\tau}_4)$$

with $p(\boldsymbol{\mu}_3, \tau_3)$ and $p(\boldsymbol{\mu}_4, \tau_4)$ the prior distributions. Thus, $f_1, ..., f_{J+1}$ are obtained through the Gibbs kernel, while f_{J+2} is obtained using the MH algorithm (Nandram 1998). We obtain the latent variables z_{ij} through one of the condi tional posterior densities of the MH algorithm. A sketch of the procedure is given in Appendix 1.

We drew 5,500 iterates, threw out the first 500, and took every fifth (obtained by trace plots). This strategy was satisfactory to wash out the autocorrelation among the iterates and to have good jumping probabilities (0.25 – 0.50) for the Metropolis steps. For the computation, first we set the hyper-parameters $\eta_1^{(0)}, \nu_1^{(0)}, \eta_{21}^{(0)}, \nu_{21}^{(0)}, \eta_3^{(0)}, \nu_3^{(0)}, \eta_{4j}^{(0)}, \nu_{4j}^{(0)}$, $j = 1, ..., J$ equal to 0. Then we ran our MH algorithm to obtain posterior samples of τ_1, τ_{21}, τ_3 and τ_{4j}, $j = 1, ..., J$. To ensure proper posterior densities, we estimate $\eta_1^{(0)}, \nu_1^{(0)}, \eta_{21}^{(0)}, \nu_{21}^{(0)}, \eta_3^{(0)}, \nu_3^{(0)}, \eta_{4j}^{(0)}, \nu_{4j}^{(0)}$ $j = 1, ..., J$, by fitting the gamma priors on the posterior samples for τ_1, τ_{21}, τ_3 and τ_{4j}, $j = 1, ..., J$. These values are shown in Table 3. Finally, with these proper priors we ran our algorithm to obtain posterior samples. Specifically, we obtained $M = 1{,}000$ iterates $(\mathbf{p}_i^{(h)}, \delta_i^{(h)})$, $h = 1, ..., M$, $i = 1, ..., c$. Inference about the $\mathbf{p}_i$, δ_i and any function of them can be made using these iterates in a straightforward manner.

Table 3
Estimates of $\eta^{(0)}$ and $\nu^{(0)}$ Corresponding to the Gamma Densities on τ_1, τ_{21} for 45+ and $\tau_3, \tau_{41}, \tau_{42}, \tau_{43}$ for 45– by Race and Sex

			Age					
			45–				45+	
Race	Sex		τ_3	τ_{41}	τ_{42}	τ_{43}	τ_1	τ_{21}
W	M	$\eta^{(0)}$	3.698	2.341	3.085	2.685	4.408	3.941
		$\nu^{(0)}$	0.036	0.071	0.201	0.163	0.009	0.052
	F	$\eta^{(0)}$	4.200	3.294	2.481	1.819	4.788	4.384
		$\nu^{(0)}$	0.030	0.059	0.072	0.017	0.008	0.019
B	M	$\eta^{(0)}$	4.948	2.922	3.156	2.404	5.971	4.376
		$\nu^{(0)}$	0.068	0.096	0.169	0.147	0.107	0.036
	F	$\eta^{(0)}$	3.745	3.084	1.893	2.350	3.292	4.488
		$\nu^{(0)}$	0.055	0.036	0.049	0.116	0.009	0.036

4. An Analysis of the NHANES III Data

In this section we illustrate our methodology using the BMI data from NHANES III. First, we study our estimates based on summary measures over the counties. Specifically, we use the weighted posterior distributions of the p_{ij},

$$\tilde{p}_j = \sum_{i=1}^{c} n_i p_{ij} \Big/ \sum_{i=1}^{c} n_i, \; j = 1, 2, 3$$

and the weighted posterior distribution of the δ_i

$$\tilde{\delta} = \sum_{i=1}^{c} n_i \delta_i \Big/ \sum_{i=1}^{c} n_i$$

for each of the age-race-sex domains. Then, for the first four examples in Table 2 we show small area effects.

We also show how to relate the p_{ijk} and the $\boldsymbol{\pi}_{ij}$ to age, race and sex using linear and nonlinear logistic regression models

4.1 Data Analysis

First, we performed a sensitivity analysis to assess the specifications of $\eta^{(0)}$ and $\nu^{(0)}$. We compared three choices of hyper-parameters $\Omega = (\eta^{(0)}, \nu^{(0)})$ to check the sensitivity of the specification of the hyper-parameters on inference. Our first choice is 4 times of Ω, *i.e.*, $4\Omega = (4\eta^{(0)}, 4\nu^{(0)})$; our second choice is the hyper-parameters without any change, *i.e.*, $\Omega = (\eta^{(0)}, \nu^{(0)})$; and our third choice is one fourth of Ω *i.e.*, $\Omega/4 = (\eta^{(0)}/4, \nu^{(0)}/4)$.

Table 4 shows the simulation results for the sensitivity to the inference of $\tilde{p}_j$ for the younger group (45–). The point estimates and standard deviations of the proportions are very similar over the three choices of hyper-parameters. Similarly, Table 5 shows the simulation results for $\tilde{p}_j$ for the older group (45+). The point estimates for males are very similar over the three choices of the hyper-parameters, but there are small changes in the point estimates for females from 4Ω to Ω. The standard deviations are increased when Ω decreases for the females, but no substantial changes are detected for males. Generally, the nonignorable nonresponse model performs better than the ignorable nonresponse model, as the nonignorable nonresponse model is not sensitive to choices of the hyper-parameters.

Table 4
Sensitivity of $\tilde{p}_j$ for Choice of $\eta_3^{(0)}, \nu_3^{(0)}, \eta_{4j}^{(0)}$ and $\nu_{4j}^{(0)}$, $j = 1, ..., 4$ for the Younger Group (45–) for the Three BMI Levels

Race	Sex	$\tilde{p}_1$	std $(\tilde{p}_1)$	$\tilde{p}_2$	std $(\tilde{p}_2)$	$\tilde{p}_3$	std $(\tilde{p}_3)$
(a) 4Ω							
W	M	0.428	0.022	0.216	0.019	0.356	0.022
	F	0.476	0.025	0.232	0.020	0.292	0.024
B	M	0.419	0.020	0.212	0.016	0.369	0.020
	F	0.434	0.026	0.185	0.023	0.381	0.027
(b) Ω							
W	M	0.427	0.022	0.211	0.020	0.362	0.025
	F	0.476	0.026	0.223	0.024	0.301	0.031
B	M	0.419	0.020	0.208	0.017	0.373	0.022
	F	0.435	0.025	0.178	0.026	0.387	0.029
(c) $\Omega/4$							
W	M	0.427	0.022	0.210	0.021	0.364	0.027
	F	0.475	0.026	0.220	0.026	0.304	0.034
B	M	0.419	0.020	0.206	0.018	0.375	0.024
	F	0.435	0.025	0.177	0.028	0.388	0.029

Note 1: $\Omega = (\eta_3^{(0)}, \nu_3^{(0)}, \eta_{41}^{(0)}, \nu_{41}^{(0)}, \eta_{42}^{(0)}, \nu_{42}^{(0)}, \eta_{43}^{(0)}, \nu_{43}^{(0)})$.
Note 2: The nonignorable nonresponse model is applied to the younger group.

Table 5
Sensitivity of $\tilde{p}_j$ for Choice of $\eta_1^{(0)}, \nu_1^{(0)}, \eta_{21}^{(0)}, \nu_{21}^{(0)}$ for the Older Group (45+) for the Three BMI Levels

Race	Sex	$\tilde{p}_1$	sdt $(\tilde{p}_1)$	$\tilde{p}_2$	std $(\tilde{p}_2)$	$\tilde{p}_3$	std $(\tilde{p}_3)$
(a)4Ω							
W	M	0.030	0.005	0.306	0.018	0.664	0.018
	F	0.081	0.002	0.436	0.004	0.483	0.004
B	M	0.053	0.011	0.317	0.017	0.630	0.018
	F	0.075	0.005	0.201	0.004	0.724	0.006
(b)Ω							
W	M	0.031	0.005	0.292	0.016	0.677	0.016
	F	0.063	0.002	0.443	0.006	0.494	0.005
B	M	0.053	0.011	0.316	0.019	0.631	0.020
	F	0.066	0.012	0.237	0.018	0.697	0.019
(c)$\Omega/4$							
W	M	0.031	0.005	0.293	0.018	0.676	0.019
	F	0.073	0.015	0.359	0.011	0.568	0.019
B	M	0.053	0.010	0.317	0.018	0.630	0.019
	F	0.065	0.013	0.221	0.022	0.714	0.025

Note 1: $\Omega = (\eta_1^{(0)}, \nu_1^{(0)}, \eta_{21}^{(0)}, \nu_{21}^{(0)})$.
Note 2: The ignorable nonresponse model is applied to the older group.

Table 6
Point Estimates and 95% Credible Intervals for the Weighted Probability of Response, $\tilde{\delta} = \sum_{i=1}^{c} n_i \delta_i / \sum_{i=1}^{c} n_i$, for Three Choices of Ω and the Younger Group

		4Ω			Ω			$\Omega/4$		
Race	Sex	$\tilde{\delta}$	std $(\tilde{\delta})$	Interval	$\tilde{\delta}$	std $(\tilde{\delta})$	Interval	$\tilde{\delta}$	std $(\tilde{\delta})$	Interval
W	M	0.775	0.016	(0.744, 0.805)	0.769	0.017	(0.735, 0.801)	0.767	0.018	(0.732, 0.799)
	F	0.855	0.017	(0.821, 0.886)	0.855	0.020	(0.810, 0.887)	0.853	0.022	(0.806, 0.887)
B	M	0.786	0.016	(0.752, 0.817)	0.780	0.018	(0.740, 0.813)	0.778	0.018	(0.739, 0.811)
	F	0.880	0.013	(0.854, 0.902)	0.878	0.015	(0.845, 0.903)	0.876	0.015	(0.838, 0.903)

Note: See the note to Table 1.

Table 6 shows point estimates of the probability of responding $\tilde{\delta}$ and their 95% credible intervals for three choices of Ω. The probabilities of responding for males are lower than those for females, and this pattern remains the same for three choices of Ω. If a similar survey is conducted in the future, we should increase the sample size by $1.30 = (1/0.769)$ times for white males and $1.17 = (1/0.855)$ times for white females (*e.g.*, if complete data are required from 1,000 households, the interviewer needs to contact 1,300 white males).

In Table 7 we present 95% credible intervals for the $\tilde{p}_j$ for the three BMI levels. For the younger group, $\tilde{p}_1$ of BMI level 1 is the highest, and $\tilde{p}_2$ of BMI level 2 is the lowest. The lower bounds for $\tilde{p}_1$ and $\tilde{p}_3$ are similar for the younger group except for white females, and those for $\tilde{p}_2$ are similar except for the non-white females. For the older group, $\tilde{p}_3$ of BMI level 3 is highest, and $\tilde{p}_1$ of BMI level 1 is lowest. Specifically $\tilde{p}_1, \tilde{p}_2$ are high and $\tilde{p}_3$ is low for the white males.

Table 7
95% Credible Intervals for the Weighted Proportions, $\tilde{p}_j = \sum_{i=1}^{c} n_i p_{ij} / \sum_{i=1}^{c} n_i$ by Age, Race and Sex

			95% credible Interval		
Age	Race	Sex	$\tilde{p}_1$	$\tilde{p}_2$	$\tilde{p}_3$
45–	W	M	(0.382, 0.470)	(0.174, 0.252)	(0.314, 0.412)
		F	(0.425, 0.525)	(0.171, 0.269)	(0.243, 0.371)
	B	M	(0.381, 0.455)	(0.176, 0.241)	(0.333, 0.419)
		F	(0.385, 0.482)	(0.130, 0.230)	(0.329, 0.442)
45+	W	M	(0.022, 0.041)	(0.255, 0.326)	(0.643, 0.710)
		F	(0.059, 0.068)	(0.431, 0.451)	(0.486, 0.505)
	B	M	(0.035, 0.076)	(0.282, 0.352)	(0.592, 0.670)
		F	(0.040, 0.093)	(0.206, 0.265)	(0.661, 0.731)

Note 1: The nonignorable nonresponse model is applied to the younger group.
Note 2: The ignorable nonresponse model is applied to the older group.

As suggested by a referee, we have looked at the results for older white females (45+) in Table 7 in greater detail. From Table 1 the observed proportions in the three BMI levels are 0.079, 0.347 and 0.568. However, the 95% credible intervals for the population proportions in Table 7 are (0.059, 0.068), (0.431, 0.451) and (0.486, 0.505) respectively. That is, while the observed proportions are close to the intervals, none of these intervals contains the observed proportions. We can explain this phenomenon in the following manner. The data for older white females (45+) are very sparse. For the 34 counties the quartiles of the observed counts in the three BMI levels are (0, 1, 3), (3, 6, 10) and (5, 9, 14) respectively. Thus, when the ignorable nonresponse model is fit to the 34 counties, there is shrinkage not only across the counties but also across the BMI levels. Consequently, the largest proportion tends to be smaller and the smallest proportion tends to be larger, and since the three proportions must add up to one, the second proportion must also "shrink" somewhat. In addition, consider the sensitivity analysis in Table 5. We can approximate 95% credible intervals for $\tilde{p}_1, \tilde{p}_2$ and $\tilde{p}_3$, by using the posterior mean $\pm 2\times$ standard deviation. The intervals at 4Ω and Ω do not contain the observed proportions, but the intervals at $\Omega/4$ do. Therefore, because of the sparseness of the data, there is some sensitivity to inference for older white females (45+) with respect to the prior misspecification of Ω. These results are expected within the small area context, when there are sparse data.

We use the first four examples in Table 2 to illustrate small area estimation. As it can be imagined, it is too cumbersome to present all the estimates for the 34 counties and the 8 domains. Table 8 shows the posterior means, standard deviations and 95% credible intervals for the p_{ij} and the δ_i.

First, we compare the estimates of the p_{ij} from the ignorable and nonignorable nonresponse models. The estimates from the two models are generally different with the intervals for the nonignorable nonresponse model wider than those for the ignorable nonresponse model.

Second, we consider the estimates (based on the non-ignorable nonresponse model) of p_{ij} for the individual counties in Table 8 with the overall averages, the $\tilde{p}_j$ in Table 7. As expected, when the $\tilde{p}_j$ are obtained, there is an overall reduction in variability because of the extra smoothing, thereby making the intervals for the smaller domains relatively much wider. In fact, all the intervals for the small domains contain the intervals for $\tilde{p}_j$.

Finally, in Table 8 we consider the estimates of $\tilde{p}_{ij}$ for the individual counties with the overall average, $\tilde{p}_j$ in Table 7. The message is similar to that for the p_{ij}.

However, we note that the first example is an exception where the credible interval for δ_i (0.459, 0.773) is almost completely to the left side of the credible interval for $\tilde{\delta}$ (0.735, 0.801). Thus, there is much shrinkage for this example which is due to the relatively large number of nonrespondents, 14 in this county for white males 45–.

Table 8
Comparaison of the Ignorable (ig) and the Nonignorable (nig) Nonresponse Models for the Four Examples (Ex) Corresponding to Small Domains Using the Cell Probabilities (p_j) and the Probability of Responding (δ)

Ex	Model		p_1	p_2	p_3	δ
1	ig	avg	0.444	0.308	0.248	
		std	0.073	0.067	0.067	
		CI	(0.297, 0.593)	(0.193, 0.450)	(0.125, 0.386)	
	nig	avg	0.450	0.276	0.273	0.637
		std	0.093	0.079	0.082	0.081
		CI	(0.256, 0.638)	(0.137, 0.444)	(0.133, 0.448)	(0.459, 0.773)
2	ig	avg	0.480	0.308	0.213	
		std	0.075	0.066	0.062	
		CI	(0.324, 0.619)	(0.193, 0.452)	(0.097, 0.344)	
	nig	avg	0.493	0.263	0.244	0.879
		std	0.074	0.065	0.062	0.041
		CI	(0.338, 0.628)	(0.141, 0.406)	(0.121, 0.394)	(0.782, 0.948)
3	ig	avg	0.420	0.306	0.274	
		std	0.071	0.063	0.063	
		CI	(0.276, 0.561)	(0.192, 0.437)	(0.161, 0.416)	
	nig	avg	0.438	0.252	0.310	0.741
		std	0.079	0.072	0.074	0.058
		CI	(0.283, 0.591)	(0.116, 0.406)	(0.186, 0.483)	(0.607, 0.836)
4	ig	avg	0.448	0.263	0.288	
		std	0.089	0.075	0.081	
		CI	(0.278, 0.620)	(0.127, 0.424)	(0.138, 0.468)	
	nig	avg	0.430	0.261	0.308	0.874
		std	0.10	0.086	0.091	0.046
		CI	(0.217, 0.619)	(0.104, 0.453)	(0.145, 0.517)	(0.768, 0.948)

Note: For each parameter avg = posterior mean; std = posterior standard deviation; CI = 95% credible interval

4.2 Linear and Nonlinear Logistic Regression Models

Let q_{ijl} denote the probability that a respondent in l^{th} $(l = 1, 8)$ age-race-sex group in the i^{th} county belongs to the j^{th} BMI level (we add the subscript l to the p_{ij} to denote the domains). Letting $v_{ijl} = \log\{\sum_{\delta=1}^{j} q_{i\delta l} / (1 - \sum_{\delta=1}^{j} q_{i\delta l})\}$, $j = 1, \ldots, J-1$, we take

$$v_{ijl} = (\theta_j - (\mu_i + \alpha_l)) / \psi_i \qquad (10)$$

subject to the constraints $\sum_{i=1}^{c} \mu_i = 0, \sum_{j=1}^{J-1} \theta_j = 0$, $\sum_{l=1}^{8} \alpha_l = 0$, and $\sum_{i=1}^{c} \ln \psi_i = 0$. The parameters $\theta_j, \mu_i, \alpha_l$ and ψ_i in (10) have posterior distributions whose properties are inherited from the posterior distributions of q_{ijl}. Each iterate of the MH algorithm provides a value for q_{ijl} which is used in (10), and a nonlinear least squares problem is solved using an iterative method to get the values of $\theta_j, \mu_i, \alpha_l$ and ψ_i (see Appendix 2). Alternatively, we can also use the much simpler linear logistic model in which the ψ_i in (10) are taken equal to unity. In this case, the least squares estimators of $\theta_j, \varphi_i, \mu_i$ and α_l exist in closed form at the h^{th} iteration of MH algorithm. Specifically, for $\varphi_i = 0$, we have the least squares estimates $\hat{\mu}_i = \bar{v}_{\ldots} - \bar{v}_{i..}$, $\hat{\theta}_j = v_{.j.}$, $\hat{\alpha}_l = v_{\ldots} - v_{..l}$, where

$$\bar{v}_{\ldots} = \sum_{i=1}^{c} \sum_{j=1}^{J-1} \sum_{l}^{8} v_{ijl} / 8c(J-1),$$

$$\bar{v}_{i..} = \sum_{j=1}^{J-1} \sum_{l=1}^{8} v_{ijl} / 8(J-1),$$

$$\bar{v}_{.j.} = \sum_{i=1}^{c} \sum_{l=1}^{8} v_{ijl} / 8c$$

and $\bar{v}_{..l} = \sum_{i=1}^{c} \sum_{j=1}^{J-1} v_{ijl} / c(J-1)$. The nonlinear least squares problem is solved using an iterative method to get the values of $\hat{\theta}_j, \hat{\varphi}_i, \hat{\mu}_i$ and $\hat{\alpha}_l$.

We present 95% credible intervals for θ_1, θ_2 and $\alpha_1, \ldots, \alpha_8$ for the younger and older groups by regression type in Table 9. For the cut-points θ_j, θ_1 gives a large negative effect compared to θ_2. The relative measure $\alpha_l (l = 1, \ldots, 4)$ of the younger group gives a negative effect, while the relative measure $\alpha_l (l = 5, \ldots, 8)$ of the older group gives positive effects. The 95% credible intervals for linear and nonlinear estimates are essentially the same.

We also relate the probability of response, $\delta_i = \sum_{j=1}^{J} \pi_{ij} p_{ij}$, to race and sex using linear and nonlinear logistic regression models for the younger group. The 95% credible intervals for θ and $\alpha_1, \ldots, \alpha_4$ for the young group by regression type are shown in Table 10. Credible intervals for all α_l for the nonlinear model are shorter than those for the linear model. However, for the nonlinear model the credible interval for θ is wider than and on the right of that for the linear model.

Table 9
Comparaison of 95% Credible Intervals for θ_1, θ_2 and $\alpha_1, \ldots, \alpha_8$ for Both Younger and Older Groups by Regression Type

	Linear	Nonlinear
θ_1	(–1.743, –1.469)	(–1.731, –1.466)
θ_2	(0.028, 0.196)	(0.025, 0.193)
α_1	(–1.167, –0.751)	(–1.159, –0.751)
α_2	(–1.395, –0.939)	(–1.385, –0.937)
α_3	(–1.127, –0.723)	(–1.119, –0.728)
α_4	(–1.112, –0.659)	(–1.103, –0.658)
α_5	(1.198, 1.514)	(1.188, 1.498)
α_6	(0.513, 0.689)	(0.506, 0.685)
α_7	(0.715, 1.210)	(0.725, 1.225)
α_8	(0.809, 1.310)	(0.803, 1.300)

Table 10
Comparaison of 95% Credible Intervals for θ and $\alpha_1, \ldots, \alpha_4$ for the Younger Group by Regression Type

	Linear	Nonlinear
θ	(1.455, 1.729)	(1.664, 2.174)
α_1	(0.165, 0.592)	(0.146, 0.523)
α_2	(−0.535, 0.014)	(−0.467, 0.007)
α_3	(0.078, 0.546)	(0.079, 0.484)
α_4	(−0.704, −0.165)	(−0.638, −0.169)

5. A Simulation Study

We describe a small simulation study to assess the performance of our multinomial nonignorable nomesponse model. We focus on the probability of responding.

We use the observed data from younger white males to obtain the posterior means of p_{i1}, p_{i2}, p_{i3} and $\pi_{i1}, \pi_{i2}, \pi_{i3}$ for each county. These are taken to be the true (t) values which we denote by $p_{i1}^{(t)}, p_{i2}^{(t)}, p_{i3}^{(t)}$ and $\pi_{i1}^{(t)}, \pi_{i2}^{(t)}, \pi_{i3}^{(t)}$. Thus, the true probability of responding in the i^{th} county is $\delta_i^{(t)} = \sum_{j=1}^{3} \boldsymbol{p}_{ij}^{(t)} \boldsymbol{\pi}_{ij}^{(t)}$ and the weighted probability of responding is $\tilde{\delta}^{(t)} = \sum_{i=1}^{c} n_i \delta_i^{(t)} / \sum_{i=1}^{c} n_i$. In our simulated examples, we used the n_i as in the BMI data for younger white males, and we kept the $p_{ij}^{(t)}$ fixed throughout. However, we varied the π_{ij} in the following manner. We kept π_{i1} fixed at $\pi_{i1}^{(t)}$, and we denote the vector of the π_{i1} by π_1. The 34 values of the $\pi_{i1}^{(t)}$ range from 0.73 to 0.83. Then, we set $\pi_2 = a\pi_1$ and $\pi_3 = b\pi_1$, where a, b = 0.8, 0.9, 1.0 (we denote the vectors of the π_{i2} and the π_{i3} by π_1 and π_2 respectively). Thus, there are 9 simulated examples.

Then, for each (a, b) we generated counts for a multinomial probability mass function with probabilities $p_{i1}^{(t)}\,\pi_{i1}$, $p_{i2}^{(t)}\,\pi_{12}$, $p_{i3}^{(t)}\,\pi_{13}$, $p_{i1}^{(t)}(1-\pi_{i1})$, $p_{i2}^{(t)}(1-\pi_{12})$, $p_{i3}^{(t)}(1-\pi_{13})$. We denote these cell counts by $y_{i1}, y_{i2}, y_{i3}, z_{i1}, z_{i2}, z_{i3}$ and the number of respondents is $r_i = \sum_{j=1}^{3} y_{ij}$. Then, we fit the nonignorable nonresponse model to the above data using the MH sampler, and we obtained M = 1,000 values $(\boldsymbol{p}_{ij}^{(h)}, \boldsymbol{\pi}_{ij}^{(h)}),\ h = 1, \ldots, M$. For each value, we computed $\tilde{\delta}^{(h)} = \sum_{i=1}^{c} n_i \delta_i^{(h)} / \sum_{i=1}^{c} n_i$ where $\delta_i^{(h)} = \sum_{j=1}^{3} p_{ij}^{(h)} \pi_{ij}^{(h)}$.

In Table 11 we report posterior means, standard deviations, numerical standard errors (using the batch means method) and 95% credible interval for the probability of responding for each choice of (a, b). We also computed $\Pr(\tilde{\delta} < \tilde{\delta}^{(t)} \mid \mathbf{y}, \mathbf{r})$ by counting the number of $\tilde{\delta}^{(h)}$ that are as large as $\tilde{\delta}^{(t)}$. An extremely large or small value of this latter quantity suggests model failure.

We plotted the estimates of the posterior densities of $\tilde{\delta}$ by choices of a and b which we obtained by using normal kernel density estimator with an optimal window width from an output analysis of the MH algorithm. The densities are an unimodal, peaked and almost symmetric. By increasing (a, b) from (0.8, 0.8) to (1.0, 1.0), the mode of the posterior densities increase.

Table 11
Characteristics of the Probability of Responding

		π_3		
π_2	stat	0.8 * π_1	0.9 * π_1	1.0 * π_1
0.8 * π_1	true	0.690	0.719	0.748
	avg	0.712	0.739	0.764
	std	0.016	0.015	0.014
	nse	0.0030	0.0031	0.0029
	CI	(0.678, 0.742)	(0.708, 0.767)	(0.734, 0.750)
	prob	0.082	0.095	0.135
0.9 * π_1	true	0.706	0.735	0.764
	avg	0.710	0.742	0.776
	std	0.017	0.016	0.014
	nse	0.0030	0.0031	0.0031
	CI	(0.673, 0.742)	(0.712, 0.769)	(0.745, 0.802)
	prob	0.377	0.303	0.210
1.0 * π_1	true	0.722	0.751	0.780
	avg	0.726	0.758	0.784
	std	0.017	0.015	0.015
	nse	0.0036	0.0036	0.0026
	CI	(0.693, 0.757)	(0.725, 0.784)	(0.750, 0.809)
	prob	0.399	0.318	0.380

Note: avg = posterior mean;
std = standard deviation;
nse = numerical standard error;
CI = 95% credible interval;
prob = $\Pr(\tilde{\delta} < \tilde{\delta}^{(t)} \mid y, r)$ the 34 values of π_1 range from 0.73 to 0.83.

In Table 11 we show that all the credible intervals contain the true values and the posterior means are close to the true value with the least discrepancy for the near ignorable nonresponse cases. The standard deviations are very similar across the nine simulated examples. Also, the numerical standard errors (nse) are small and similar for all nine simulated examples. The estimates of $\Pr(\tilde{\delta} < \tilde{\delta}^{(t)} \mid \mathbf{y}, \mathbf{r})$ range from 0.30 to 0.40, except for the most nonignorable nonresponse cases in which (a, b) = (0.8, 0.8) and (0.8, 0.9). Thus, the model does perform reasonably well.

6. Conclusion

We have described a Bayesian methodology that can be used to analyze multinomial data for small areas when there is nonignorable nonresponse. A hierarchical model is used, and we have shown that it performs reasonably well. In fact, we have extended the method of Stasny (1991) in two directions: (a) we have considered multinomial data with more than two cells (binomial) and (b) we have done a full Bayesian analysis. Both (a) and (b) have been implemented for small areas.

The Markov chain Monte Carlo method permits an assessment of the complex structure of the multinomial nonresponse estimation. Our empirical analysis and simulation study indicate good performance of the model for these data. Thus, the method of ratio estimation currently used in NHANES III may be replaced by our Bayesian method as the nonrespondents' characteristics might differ from those of the respondents. In fact, an application of our model to the NHANES III data shows that in each county there are substantial differences in the proportions of individuals at the three BMI levels by age and sex. This can be seen in Table 1 when the observed counts are summed over the counties. But, we have obtained inference (including measure of precision) for each county by age, race and sex.

Our methodology can be extended in three ways. First, it is feasible to use a model that incorporates an extent of nonignorability, rather than just the dichotomy of ignorable nonresponse and nonignorable nonresponse. Second, one can use other prior distributions (*e.g.*, Dirichlet process prior) to model heterogeneity in the clustering of the areas rather than assuming homogeneity of the areas as we have done. Third, one can use a fourth stage in our model to accommodate clustering within households as well as clustering within areas (counties) in NHANES III. These tasks are very difficult.

Acknowledgement

This work was done at the National Center For Health Statistics while Balgobin Nandram was the first ASA/NCHS Research Fellow and Geunshik Han was on sabbatical leave from Hanshin University, Korea.

Appendix 1

Metropolis-Hastings Samplers

For the ignorable nonresponse model, $(\boldsymbol{\mu}_1, \tau_1)$ and (μ_{21}, τ_{21}) are independent a posteriori with

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

and

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

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

For the nonignorable nonresponse model, it is convenient to condition on **z** to obtain

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

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

where $p(\boldsymbol{\mu}_3, \tau_3), p(\mu_{4j}, \tau_{4j}), j = 1, ..., J$ are the prior distributions. Given **z**, (A.3) and (A.4) are independent with

$$p(z_{i1} = t_{i1}, ..., z_{iJ} = t_{ij} | \mathbf{y}, \mathbf{r}, \mu_4, \tau_4, \mu_{3j}, \tau_{3j}, j = 1, ..., J) = w_{it_{i1}t_{i2}...t_{iJ}} \Big/ \sum_{t_{i1}=0}^{n_i - r_i} \cdots \sum_{t_{iJ}=0}^{n_i - r_i} w_{it_{i1}t_{i2}...t_{iJ}}, \quad (A.5)$$

for $t_{ij} = 0, 1, ..., n_i - r_i, \sum_{j=1}^{J} t_{ij} = n_i - r_i$,

$$w_{it_{i1}t_{i2}...t_{iJ}} = \binom{n_i - r_i}{t_{i1}, ..., t_{ij}} D(\mathbf{y}_i + \mathbf{t}_i + \boldsymbol{\mu}_3 \tau_3)$$

$$\prod_{j=1}^{J} B(y_{ij} + \mu_{4j}\tau_{4j}, t_{ij} + (1-\mu_{4j})\tau_{4j}).$$

We ran the MH sampler by drawing a random deviate from each of (A.3), (A.4), and (A.5). It is easy to draw a random deviate from (A.5). Samples were obtained from each of (A.3), (A.4) and (A.5) using the MH algorithm of Nandram (1998).

Appendix 2

Nonlinear Least Squares Estimates

Let

$$v_{ijl} = \log\left\{ \sum_{s=1}^{j} q_{isl} \Big/ \left(1 - \sum_{s=1}^{j} q_{isl}\right) \right\}, j = 1, ..., J - 1 = J'.$$

These v_{ijl} are obtained for each iterate from the Metropolis-Hastings sampler. To solve the nonlinear least squares problem we minimized

$$\sum_{i=1}^{c} \sum_{j=1}^{J'} \sum_{l=1}^{8} \left\{ v_{ijl} - e^{\varphi_i} (\theta_j - (\mu_i + \alpha_l)) \right\}^2 \quad (A.1)$$

subject to the constraints $\sum_{i=1}^{c} \boldsymbol{\mu}_i = 0$, $\sum_{j=1}^{J'} \theta_j = 0$, $\sum_{l=1}^{8} \alpha_k = 0$, and letting $e^{\varphi_i} = \psi_i^{-1}, \sum_{i=1}^{c} \ln \psi_i = 0$.

Taking partial derivatives to find the least squares estimate, we have

$$\hat{\varphi}_i = \log \left\{ \frac{\sum_{j=1}^{J'} \sum_{l=1}^{8} v_{ijl} (\hat{\theta}_j - \hat{\mu}_i - \hat{\alpha})}{\sum_{j=1}^{J'} \sum_{l=1}^{8} (\hat{\theta}_j - \hat{\mu}_i - \hat{\alpha}_l)^2} \right\} = \log \psi_i^{-1} \quad (A.2)$$

where

$$\hat{\theta}_j = \left[\sum_{i=1}^{c} e^{2\hat{\varphi}_i} \left\{ \frac{1}{8} \sum_{t=1}^{8} (e^{-\hat{\varphi}_i} v_{ijl} + \hat{\mu}_i + \hat{\alpha}_l) \right\} \right] \Big/ \sum_{i=1}^{c} e^{2\hat{\varphi}_i} , \quad (A.3)$$

$$\hat{\mu}_i = \left(\frac{1}{8J'} \right) \sum_{l=1}^{8} \sum_{j=1}^{J'} \left\{ \hat{\theta}_j - (\hat{\alpha}_l + e^{-\hat{\varphi}_i} v_{ijl}) \right\} \quad (A.4)$$

and

$$\hat{\alpha}_l = \sum_{i=1}^{c} \frac{1}{J'} \sum_{j=1}^{J'} e^{2\hat{\varphi}_i} \left\{ \hat{\theta}_j - (\hat{\mu}_i + e^{-\hat{\varphi}_i} v_{ijl}) \right\} \Big/ \sum_{i=1}^{c} e^{2\hat{\varphi}_i} . \quad (A.5)$$

With these settings we draw the q_{ijl} from a MH algorithm, and the nonlinear least squares problem is solved using an iterative method to get values of φ_i, θ_j, μ_i and α_l. Let

$$v_{ijl}^{(h)} = \log \left\{ \sum_{s=1}^{j} q_{isl}^{(h)} \Big/ \left(1 - \sum_{s=1}^{j} q_{isl}^{(h)} \right) \right\},$$

where $q_{isl}^{(h)}$ denotes the value of q_{isl} at the h^{th} iterate of the MH algorithm. Then we minimize (A.l) subject to the above constraints at the h^{th} iterate to obtain $\varphi_i^{(h)}$, $\theta_j^{(h)}$, $\mu_i^{(h)}$ and $\alpha_l^{(h)}$. These iterates provide an estimate of the posterior distributions of φ_i, θ_j, μ_i and α_l. Convergence occurred for our application in less then 10 iterations.

References

Albert, J.H., and Gupta, A.K. (1985). Bayesian methods for binomial data with applications to a nonresponse problem. *Journal of the American statistical Association*, 80, 167-174.

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

Basu, D., and Pereira, C.A. (1982). On the Bayesian analysis of categorical data: The Problem of nonresponse. *Journal of Statistical Planning and lnference*, 6, 345-362.

Crawford, S. L., Johnson, W.G. and Laird, N.M. (1993). Bayes analysis of model-based methods for nonignorable nonresponse in the Harvard Medical Practice Survey (with discussions). ln *Case Studies in Bayesian Statistics* (Eds. C. Gatsonis, J.S. Hodges, R.E. Kass and N.D. Sinpurwalla). New York: Springer-Verlag, 78-117.

De Heer, W. (1999). International response trends: Results of an international survey. *Journal of Official Statistics*, 15, 129-142.

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

Forster, J.J., and Smith, 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.

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

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

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

Kaufman, G.M., and King, B. (1973). A Bayesian analysis of nonresponse in dichotomous processes. *Journal of the American Statistical Association*, 68, 670-678.

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

Malec, D., Davis, W. and Cao, X. (1999). Model-based small area estimates of over-weight prevalence using sample selection adjustment. *Statistics in Medicine*, 18, 3189-3200.

Mohadjer, L., Bell, B. and Waksberg, J. (1994). National Health and Nutrition Examination Survey III-accounting for item nonresponse bias. *National Center for Health Statistics*.

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

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

National Center For Health Statistics (1994). Plan and operation of the Third National Health and Nutrition Examination Survey. *Vital and Health Statistics*, Series 1, 32.

Olson, R.L. (1980). A least squares correction for selectivity bias. *Econometrica*, 48, 1815-1820

Park, T. (1998). An approach to categorical data nonignorable nonresponse. *Biometrics*, 54, 1579-1590.

Park, T., and Brown, M.B. (1994). Models for categorical data with nonignorable non-response. *Journal of the American Statistical Association*, 89, 44-52.

Rao, C.R. (1973). *Linear Statistical Inference and Its Applications*. New York: John Wiley & Sons, Inc.

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

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

Schafer, J.L., Ezzati-Rice, T.M., Johnson, W., Khare, M., Little, R.J.A. and Rubin, D.B. (1996). The NHANES III multiple imputation project. Survey Research Methods, *Proceedings of the American Statistical Association*, 28-37.

Stasny, E.A. (1991). Hierarchical models for the probabilities of a survey classification and nonresponse: An example from the National Crime Survey. *Journal of the American Statistical Association*, 86, 296-303.

Stasny, E.A., Kadane, J.B. and Fritsch, K.S. (1998). On the fairness of death penalty jurors: A comparison of Bayesian models with different levels of hierarchy and various missing-data mechanisms. *Journal of the American Statistical Association*, 93, 464-477.