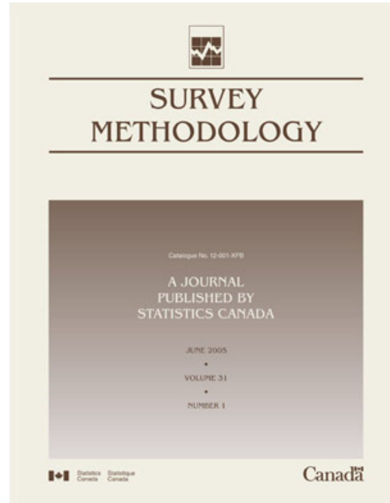
Catalogue no. 12-001-XIE

Survey Methodology

June 2005



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

June 2005

Published by authority of the Minister responsible for Statistics Canada

May 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. 31, No. 1, pp. 73-84
Statistics Canada, Catalogue No. 12-001

Hierarchical Bayesian Nonignorable Nonresponse Regression Models for Small Areas: An Application to the NHANES Data

Balgobin Nandram and Jai Won Choi[1]

Abstract

We use hierarchical Bayesian models to analyze body mass index (BMI) data of children and adolescents with nonignorable nonresponse from the Third National Health and Nutrition Examination Survey (NHANES III). Our objective is to predict the finite population mean BMI and the proportion of respondents for domains formed by age, race and sex (covariates in the regression models) in each of thirty five large counties, accounting for the nonrespondents. Markov chain Monte Carlo methods are used to fit the models (two selection and two pattern mixture) to the NHANES III BMI data. Using a deviance measure and a cross-validation study, we show that the nonignorable selection model is the best among the four models. We also show that inference about BMI is not too sensitive to the model choice. An improvement is obtained by including a spline regression into the selection model to reflect changes in the relationship between BMI and age.

Key Words: Cross-validation; Deviance; Metropolis-Hastings sampler; Normal-logistic regression model; Spline regression model.

1. Introduction

The National Health and Nutrition Examination Survey (NHANES III) is one of the surveys used by the National Center for Health Statistics (NCHS) to assess the health of the U.S. population. One of the variables in this survey is body mass index (BMI), and the World Health Organization has used BMI to define overweight and obesity. Under ignorability estimators from the NHANES III data are biased because there are many nonrespondents, and the main issue we address here is that nonresponse should not be ignored because respondents and nonrespondents may differ. The purpose of this work is to predict the finite population mean BMI for children and adolescents, post-stratified by county for each domain formed by age, race and sex and to investigate what adjustment needs to be made for nonignorable nonresponse. Our approach is to fit several hierarchical Bayesian models to accommodate the nonresponse mechanism.

Recently, several articles have been written about overweight and obesity. In outlining the first national plan of action for overweight and obesity, the Surgeon General called for sweeping changes in schools, restaurants, workplaces and communities to help combat the growing epidemic of Americans who are overweight or obese. He said that the obesity report “Is not about esthetics and it’s not about appearances. We’re talking about health.” As noted by Squires (2001) “Health care costs for overweight and obesity total an estimated $117 billion annually.” Overweight children often become overweight in adulthood, and overweight in adulthood is a health risk (Wright, Parker, Lamont and Craft 2001). In a very interesting article, using NHANES data Ogden, Flegal, Carroll and Johnson (2002) describe the most recent national estimates of the prevalence and trends in overweight among U.S. children and adolescents. Based on a limited analysis they conclude “The prevalence of overweight among children in the United States is continuing to increase especially among Mexican-American and non-Hispanic black adolescents.” Several disorders have been linked to overweight in childhood. A potential increase in type 2 diabetes mellitus is related to the increase in overweight among children (Fagot-Campagna 2000); so are cardiovascular risk factor, high cholesterol levels, and abnormal glucose levels (Dietz 1998). Thus, it would be helpful to study the BMIs for children and adolescents using methods that can provide accurate adjustment for nonresponse and better measure of precision.

Letting x denote covariates and y the response variable, Rubin (1987) and Little and Rubin (1987) 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 and (c) depends on the y and possibly x. The missing data are missing completely at random (MCAR) in (a), missing at random (MAR) in (b) and one may say that 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 are distinct (Rubin 1976). Models for MNAR missing-data mechanisms are called nonignorable.

1. Balgobin Nandram, Department of Mathematical Sciences, Worcester Polytechnic Institute, 100 Institute Road, Worcester, MA 01609-2280. E-mail: balnan@wpi.edu; Jai Won Choi, National Center for Health Statistics, 3311 Toledo Road, Hyattsville, MD 20782. E-mail: jwc7@cdc.gov.

Nonresponse models can be classified very broadly into selection and pattern mixture models (*e.g.*, see Little and Rubin 1987). Let $[y]$ and $[r]$ denote respectively the density function of the response variable y, and the response indicator r, with obvious notations for the joint and conditional densities. Then the selection model specifies that $[y,r]=[r\mid y][y]$ and the pattern mixture model specifies $[y,r]=[y\mid r][r]$. The selection approach was developed to study sample selection problems (*e.g.*, Heckman 1976 and Olson 1980). While the two models have the same joint density, in practice the components $[r\mid y]$ and $[y]$ for the selection model, and $[y\mid r]$ and $[r]$ for the pattern mixture model are specified. Thus, these models may differ.

Thus, we use two nonignorable nonresponse models, a selection model and a pattern mixture model, to analyze the NHANES III data. Each model is used in the hierarchical Bayesian frame work for our nonignorable nonresponse problem, and to study sensitivity to model choice the results are compared. In the selection model, the response propensity is related to BMI only, and then the model on BMI has a linear model on age, race, sex and the interaction of race and sex. In the pattern mixture model, the propensity to respond is related to age, race and sex (not BMI), and the model on BMI has two closely related linear forms on age, race, sex and the interaction of race and sex. These two models hold for the entire population. The BMI values of the nonrespondents and the nonsampled individuals are predicted from each model. We prefer the selection model because we can incorporate the structure in the NHANES III data, and based on statistical arguments this turns out to be true.

Greenlees, Reece and Zieschang (1982) developed a normal-logistic regression model for imputing missing values when the probability of response depends upon the variable being imputed. They applied the model to data on wages and salary in the Current Population Survey (CPS) data on wages. David, Little, Samuel and Triest (1986) compared the CPS hot deck method and the normal-logistic regression model to wages and salary from a similar data set, and they found very little difference between the two methods. We note that the normal-logistic regression model is a nonignorable nonresponse selection model, but it does not account for clustering. To accommodate clustering within counties in the NHANES III data, it is natural to start with the normal-logistic model.

Our hierarchical Bayesian selection model has a special structure. In NHANES III the propensity to respond increases with age (race and sex play a minor role), and doctors believe that obese individuals tend not to turn up for the physical examination. Thus, given the BMI values, like Greenlees *et al*. (1982) the response indicators follow a logistic regression model with the logarithm of the BMI values being the covariate. In turn, the logarithms of the BMI values are distributed according to a linear model in which the covariates are age, race and sex. This is the most important information we incorporate into the selection model. In addition, unlike Greenlees *et al.* (1982) our model includes clustering effects to account for heterogeneity among counties through the response indicators and the BMI values. Here each county has its own set of parameters, and there is a common distribution over these sets of parameters. This is also an important prior information we incorporate into our model, and it is one of the attractive features of the hierarchical Bayesian methodology.

In the Bayesian approach, the main difficulty is formulating the relationship between the respondents and nonrespondents. This latter issue can be accommodated within the selection approach through the normal-logistic structure. We also consider a hierarchical Bayes model within the pattern mixture approach. The pattern mixture model is a useful alternative to study sensitivity to the assumption in the selection model. To assess the assumption of nonignorable nonresponse, we also consider special cases of the selection and pattern mixture models to obtain two ignorable models. We found that a fifth model is required, in which we extend our selection model to a spline regression model to accommodate the dynamic relation between BMI and age.

Nandram, Han and Choi (2002) developed a methodology to analyze the BMI data by age, race and sex when BMI is categorized into three intervals. This is a multinomial extension of the nonresponse nonignorable analysis of Stasny (1991) for binary data. This methodology applies generally to any number of cells in several areas (counties in our application). Nandram and Choi (2002 a,b) consider further extensions of the work of Stasny for binary data (*i.e.*, data from the National Health Interview Survey and the National crime survey). Here we do not categorize the BMI values, but rather we treat them in their own right as continuous values. The quantities of interest are the finite population mean BMI and the proportion of responding individuals in each domain formed by age, race, sex and county.

The rest of the paper is organized as follows. In section 2, we briefly describe the NHANES III data. In section 3, we discuss the hierarchical Bayesian models for ignorable and nonignorable nonresponse. We also describe the model fitting, model selection and assessment which use predictive deviance and cross-validation. In section 4 we describe the analysis of NHANES III BMI data. Section 5 has a description of a spline regression model and comparisons. Finally, section 6 has concluding remarks about our approach.

2. 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. The number of sampled individuals in each age-race-sex group is known for each county. The sample size by county, age, race and sex are relatively sparse. 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 counties across the continental United States during the period from October 1988 through September 1994, but for confidentiality reasons the final data of this study came from only the 35 largest counties (from 14 states) with population at least 500,000 for selected age categories by sex and race. In this paper, we analyze public use data from these 35 counties; the demographic variables are age, race and sex, and the health indicator of our interest is body mass index (BMI), weight in kilograms divided by the square of height in meters (Kuczmarski, Carrol, Flegal and Troiano 1997). The World Health Organization (WHO Consultation of Obesity 2000) has designated an adult with BMI at least 30 as obese; overweight refers to adults with BMI in the range [25, 30). For children $1-6$ years old and adolescents 7–19 years old overweight and obesity are age-dependent.

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. Here we do not consider the small number of individuals whose BMI values and covariates (age, race and sex) are missing (*i.e.*, unit nonresponse). For simplicity and for all practical purposes it is reasonable to include all individuals with their covariates (*i.e.*, complete data and item nonresponse) reported in our data analysis. 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." We note also that for children and adolescents the observed nonresponse rate is about 24%. A partial reason for the nonresponse for young children is that the parents or older mothers were extremely protective and would not allow their children to leave home for a physical examination.

We study the BMI data for four age classes (02 – 04, 05 – 09, 10 – 14, 15 – 19 years). Recalling that there are 560 ($35 \times 4 \times 2 \times 2$) domains, the sample sizes on the average are very small per domain (*e.g.*, 2,647/560 $\approx$ 4). Thus, there is a need to "borrow strength" from the domains. Also, the sample size is small relative to the finite population size (*e.g.*, 100 $\times$ (2,647/6,653,738) = 0.04%). The prediction problem needs much computation. The observed data indicate that there is an increasing trend of BMI with age with slightly increasing variability.

NHANES III data are adjusted by multiple stages of ratio weightings to be consistent with the population; see Mohadjar, Bell and Waksberg (1994). In this ratio-method, item nonresponse adjustment is done by ratio estimation within the same adjustment class and the distributions of the respondents and nonrespondents are assumed to be same. There is a need to consider methods for handling non-ignorable nonresponse other than the ratio-adjustment method. Here we present a Bayesian method as a possible alternative for studying NHANES III nonresponse.

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 in order to provide several data sets for public use. As one of the limitations of the project they stated "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. Unlike Schafer *et al.* (1996), we include clustering at the county level, although there may be 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.

For our current application, inference is required for each age, race and sex domain within county. One standard small area estimation method is to identify each small area by a parameter, and then assume a common stochastic process over the 560 parameters. But because of the sparseness of the data, this is not desirable. Thus, our models are constructed at county level, and at the same time age, race and sex are represented as covariates. Inference is made for each domain formed by crossing age, race and sex within county through our regression models. This is a key point in our analysis.

3. Hierarchical Bayesian Methodology

In this section we describe two Bayesian models for non-ignorable nonresponse, and we deduce two additional ignorable models as special cases. We describe the model selection and assessment for the selected model (*i.e.*, the selection model).

There are data from $\ell = 35$ counties and each county has N_i (known) individuals. We assume a probability sample of n_i individuals is taken from the i^{th} county. Let s denote the set of sampled units and ns the set of nonsampled units. Let r_{ij} for $i = 1, 2, \ldots, \ell$ and $j = 1, 2, \ldots, N_i$ be the response indicator ($r_{ij} = 1$ for respondents and $r_{ij} = 0$ for non-respondents) for the j^{th} individual within the i^{th} county in the population. Also, let x_{ij} be the logarithm of the BMI value. We found that the logarithm transformation gives a better representation, and we use it throughout. Note that r_{ij} and x_{ij} are all observed in the sample s but they are unknown in ns. Let $r_i = \sum_{j=1}^{n_i} r_{ij}$ (*i.e.*, r_i is the number of sampled individuals that responded in the i^{th} county).

For convenience, we express the BMI x_{ij} as $x_{i1}, x_{i2}, \ldots, x_{ir_i}, x_{ir_i+1}, \ldots, x_{in_i}$ in s and $x_{in_i+1}, \ldots, x_{iN_i}$ in ns for county i. A key point that we note for what follows is that the r_i individuals are not necessarily random respondents from the n_i individuals randomly sampled. This is the nonresponse bias we need to address. It is clear that we need to predict the BMI value x_{ij} for (a) the nonrespondents in s and (b) the individuals in ns. Thus, for the finite population of N_i individuals, we need a Bayesian predictive inference for

$$\bar{X}_i = \frac{\sum_{j=1}^{N_i} x_{ij}}{N_i} \quad \text{and} \quad P_i = \frac{\sum_{j=1}^{N_i} r_{ij}}{N_i},$$

for $i = 1, \ldots, \ell$.

Letting $\bar{x}_i^{(s,r)} = \sum_{j=1}^{r_i} x_{ij} / r_i, \bar{x}_i^{(s,nr)} = \sum_{j=r_i+1}^{n_i} x_{ij} / (n_i - r_i)$ and $\bar{x}_i^{(ns)} = \sum_{j=n_i+1}^{N_i} x_{ij} / N_i - n_i)$, we note that

$$\bar{X}_i = f_i \left\{ g_i^{(s)} \bar{x}_i^{(s,r)} + (1 - g_i^{(s)}) \bar{x}_i^{(s,nr)} \right\} + (1 - f_i) \bar{x}_i^{(ns)} \quad (1)$$

where $f_i = n_i / N_i$ and $g_i^{(s)} = r_i / n_i$. Note that while the f_i are fixed by design, the g_i and $\bar{x}_i^{(s,r)}$ are observed. Also, letting $\hat{p}_i^{(s)} = r_i / N_i$ and $\hat{p}_i^{(ns)} = (\sum_{j=n_i+1}^{N_i} r_{ij}) / (N_i - n_i)$,

$$P_i = f_i \hat{p}_i^{(s)} + (1 - f_i) \hat{p}_i^{(ns)}, \quad (2)$$

$i = 1, \ldots, \ell$. We develop our hierarchical Bayesian models to perform predictive inference for quantities like (1) and (2) depending on the domain.

3.1 Competing Models

Our models have two parts, one part for the response mechanism and the other part for the distribution of BMI. These two parts are connected to form a single model under nonignorable nonresponse or ignorable nonresponse.

First, we describe the selection model. For Part 1 of this model the response depends on the BMI as follows

$$r_{ij} \mid x_{ij}, \boldsymbol{\beta}_i \sim \text{Bernoulli} \left\{ \frac{e^{\beta_{0i} + \beta_{1i} x_{ij}}}{1 + e^{\beta_{0i} + \beta_{1i} x_{ij}}} \right\}, \quad (3)$$

$$(\beta_{0i}, \beta_{1i}) \mid \theta_0, \theta_1, \sigma_1^2, \sigma_1^2, \rho_1 \overset{\text{iid}}{\sim} \text{BVNormal}(\theta_0, \theta_1; \sigma_1^2, \sigma_2^2, \rho_1), \quad (4)$$

$$\boldsymbol{\theta} \sim N(\boldsymbol{\theta}^{(0)}, \Delta^{(0)}), \sigma_1^{-2}, \sigma_2^{-2} \sim \text{Gamma}(a/2, a/2) \quad \text{and} \quad \rho_1 \sim \text{Uniform}(-1, 1), \quad (5)$$

where $a, \boldsymbol{\theta}^{(0)}$ and $\Delta^{(0)}$ are to be specified. Note that the prior densities in (5) are all jointly independent. The assumption (3) is important because it relates the response propensity to the BMI values; doctors believe that over-weight and obese individuals tend not to come to the MECs for the examinations. Clustering among the counties is accommodated by (4), and it is this assumption that permits a "borrowing of strength" among the counties.

The second part of the model is about the BMI. The single most important predictor of BMI is age, with race and sex playing a relatively minor role. One possibility is to take the BMI values to be

$$x_{ij} = \mu_{ij} + \epsilon_{ij}, \quad \mu_{ij} = \alpha_{0ij} + \alpha_{1ij} a_{ij}$$

where a_{ij} denotes age and $\epsilon_{ij} \mid \sigma_3^2 \overset{\text{iid}}{\sim} \text{Normal}(0, \sigma_3^2)$ for $i = 1, \ldots, \ell$ and $j = 1, \ldots, N_i$. Also, there is a need to understand the relationship between BMI and age, race and sex. We let $z_{ij0} = 1$ for an intercept, $z_{ij1} = 1$ for non-black and $z_{ij1} = 0$ for black, $z_{ij2} = 1$ for male and $z_{ij2} = 0$ for female, $z_{ij3} = z_{ij1} z_{ij2}$ for the interaction between race and sex, and we let $\mathbf{z}'_{ij} = (z_{ij0}, z_{ij1}, z_{ij2}, z_{ij3})$. Then, for a regression of BMI on age adjusting for race and sex, letting $\boldsymbol{\alpha}'_1 = (\alpha_{01}, \alpha_{02}, \alpha_{03}, \alpha_{04})$ and $\boldsymbol{\alpha}'_2 = (\alpha_{11}, \alpha_{12}, \alpha_{13}, \alpha_{14})$, we take $\alpha_{0ij} = \mathbf{z}'_{ij} \boldsymbol{\alpha}_1 + v_{0i}$ and $\alpha_{1ij} = \mathbf{z}'_{ij} \boldsymbol{\alpha}_2 + v_{1i}$ to get

$$\mu_{ij} = (\mathbf{z}'_{ij} \boldsymbol{\alpha}_1 + v_{0i}) + (\mathbf{z}'_{ij} \boldsymbol{\alpha}_2 + v_{1i}) a_{ij}$$

where v_{0i} and v_{1i} are random effects centered at zero with bivariate normal distribution shown below for each model.

Thus, in Part 2 of the selection model, we assume

$$x_{ij} = (\mathbf{z}'_{ij} \boldsymbol{\alpha}_1 + v_{0i}) + (\mathbf{z}'_{ij} \boldsymbol{\alpha}_2 + v_{1i}) a_{ij} + e_{ij} \quad \text{and} \quad e_{ij} \mid \sigma_3^2 \overset{\text{iid}}{\sim} \text{Normal}(0, \sigma_3^2), \quad (6)$$

$$(v_{0i}, v_{1i}) \mid \sigma_4^2, \sigma_5^2, \rho_2 \overset{\text{iid}}{\sim} \text{BVNormal}(0, 0; \sigma_4^2, \sigma_5^2, \rho_2). \quad (7)$$

Again, clustering among the counties is accommodated by (7), and it is this assumption that permits a "borrowing of strength" among the counties. For this part of the model, we use the prior

$$\boldsymbol{\alpha}_1 \sim \text{Normal}(\boldsymbol{\alpha}_2^{(0)}, \Delta_2^{(0)}) \text{ and } \boldsymbol{\alpha}_2 \sim \text{Normal}(\boldsymbol{\alpha}_3^{(0)}, \Delta_3^{(0)}),$$
$$\sigma_3^{-2}, \sigma_4^{-2}, \sigma_5^{-2}, \overset{\text{iid}}{\sim} \text{Gamma}(a/2, a/2) \quad \text{and}$$
$$\rho_2 \overset{\text{iid}}{\sim} \text{Uniform}(-1, 1) \quad (8)$$

where $a, \boldsymbol{\alpha}_k^{(0)}$ and $\Delta_k^{(0)}, k = 1, 2$ are to be specified. Note that the prior densities in (8) are all jointly independent.

The nonignorable nonresponse pattern mixture model is presented in Appendix A. We have included race, sex and their interaction in the response part of the model, although these turn out to be unnecessary. The difference between the respondents and the nonrespondents in the pattern mixture model is that the intercepts in the regression vary with counties for the respondents but not for the nonrespondents; other parameters are the same. In this way we are able to "center" the nonignorable nonresponse model on the ignorable nonresponse model with some variation; see Nandram and Choi (2002 a) for a similar idea. We need to do so because the parameters become unidentifiable if substantial difference between the respondents and the nonrespondents is assumed in the nonignorable nonresponse model without the scientific knowledge. While we have used random effects to discriminate between the respondents and the nonrespondents, the parameters providing systematic difference between the respondents and nonrespondents in model of Rubin (1977), are not identifiable. Note that while in the pattern mixture model in (A.4) there are two specifications/patterns for x_{ij} (*i.e.*, $r_{ij} = 0$ and $r_{ij} = 1$), but in the selection model there is a single specification.

We show how to specify parameters like $\boldsymbol{\theta}^{(0)}, \Delta^{(0)}$, $\boldsymbol{\alpha}_k^{(0)}, \Delta_k^{(0)}, k = 1, 2$ in Appendix C. For a proper diffuse prior we choose *a* to be a value like 0.002. One can also use a shrinkage prior on σ_1^{-2} and σ_2^{-2} (see Natarajan and Kass 2000; and Daniels 1999). But this is not necessary in the hierarchical model.

It is an attractive property of the hierarchical Bayesian model that it introduces correlation among the variables. For example, in the selection model, (4) and (7) introduce a correlation among the r_{ij} and the x_{ij}, respectively. 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; this is intimately connected to the cluster effect. Yet another benefit is that there is robustness in the model specifications at deeper levels beyond the sampling process (*e.g.*, inference with (5) and (8) is fairly robust to moderate perturbations of the specifications of the hyperparameters). We have found this empirically here and elsewhere.

We obtain an ignorable nonresponse selection model by setting $\beta_{1i} = 0$ for all counties with appropriate adjustment in the selection model. For an ignorable nonresponse pattern mixture model we set $x_{ij} = (\mathbf{z}'_{ij} \boldsymbol{\alpha}_1 + v_{0i}) + (\mathbf{z}'_{ij} \boldsymbol{\alpha}_2 + v_{1i}) a_{ij} + \in_{ij}$ for both values of the r_{ij}.

3.2 Model Fitting

In this section we describe how to use the Metropolis-Hastings sampler to fit the models. We also use a deviance measure to select the best model among our four models. Then, we use a cross-validation analysis to assess the goodness of fit of the selected model, and because the same general principle applies to the four models, we describe model fitting for the selection model only.

Thus, we now combine the model for the response mechanism and the model for the BMI values to obtain the joint posterior density of all the parameters. The x_{ij} for $j = r_i + 1, \ldots, n_i$, $i = 1, \ldots, \ell$ are unknown; that is, they are latent variables. We denote these latent variables by $\boldsymbol{x}^{(s,nr)}$ and the observed data are denoted by $\boldsymbol{x}^{\text{obs}}$. Using Bayes' theorem to combine the likelihood function and joint prior distribution, we obtain the joint posterior density which, apart from the normalization constant, is $p(\boldsymbol{x}^{(s,nr)}, \boldsymbol{\sigma}^2, \boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{v}, \boldsymbol{\theta}, \rho_1, \rho_2 \mid \boldsymbol{x}^{(s,r)})$ and is given in (B.1) in Appendix B.

The posterior density in (B.1) is complex, so we used Markov chain Monte Carlo (MCMC) methods to draw samples from it. Specifically, we used the Metropolis-Hastings sampler (see Chib and Greenberg 1995 for a pedagogical discussion). We also used the trace plots and autocorrelation diagnostics reviewed by Cowles and Carlin (1996) to study convergence and we used the suggestion of Gelman, Roberts and Gilks (1996) to monitor the jumping probability in each Metropolis step in our algorithm. In performing the computation, centering the BMI values help in achieving convergence (see Gelfand, Sahu and Carlin 1995). However, this is not quite a straightforward task because centering in the logistic regression affects the BMI part of the model as well.

We obtained a sample of 1,000 iterates which we used for inference and model checking. By using the trace plots we "burn in" 1,000 iterates, and to nullify the effect of autocorrelations, we picked every tenth iterate thereafter. This rule was obtained by trial and error while tuning the Metropolis steps. We maintain the jumping probabilities in (0.25, 0.50); see Gelman *et al.* (1996).

3.3 Model Selection and Model Assessment

We used the minimum posterior predictive loss approach (Gelfand and Ghosh 1998) to select the best model among the first four.

Under squared error loss the minimum posterior predictive loss is

$$D_k = P + \frac{k}{k+1} G$$

$$P = \sum_{ij} \mathrm{Var}(x_{ij}^{\text{pre}} \mid \boldsymbol{x}^{\text{obs}}), \quad G = \sum_{ij} \left\{E(x_{ij}^{\text{pre}} \mid \boldsymbol{x}^{\text{obs}}) - x_{ij}^{\text{obs}}\right\}^2$$

where $f(x_{ij}^{\text{pre}} \mid \boldsymbol{x}^{\text{obs}}) = \int f(x_{ij}^{\text{pre}} \mid \Omega)\pi(\Omega \mid \boldsymbol{x}^{\text{obs}}) d\Omega$ and x_{ij}^{pre} are the predicted values and Ω is the set of all parameters. This measure extends the one obtained earlier (Laud and Ibrahim 1995), and we have taken $k = 100$ to match this earlier version. Note that for the nonresponse application, these measures are computed only on the complete BMI data after fitting our nonresponse models.

In Table 1 we present the deviance measure (D_{100}) and its associated components, goodness of fit (G) and the penalty (P) for the four models. Using the deviance measure the selection model is much better. While P is roughly the same, G is much smaller, making D_{100} smaller for the selection model. The difference between the two pattern mixture models are more pronounced than the difference between the two selection models. However, because standard errors are not available, it is difficult to tell the strength of the difference.

Table 1
Comparison of the Ignorable, Pattern Mixture and the Selection Models Using the Deviance Measure

Model	G	P	D_{100}
SEI	135	135	270
SE	118	135	253
PMI	268	135	403
PM	204	135	339

Note: $D_{100} = G + (100/(100+1))\,P$ where G is a goodness of fit, P a penalty and D the deviance; the pattern mixture (PM) model and the selection model (SE) are both nonignorable. SEI is ignorable version of the selection model, PMI is ignorable version of the pattern mixture model.

Next, we look for deficiencies in the selection model. We use a Bayesian cross-validation analysis to assess the goodness of fit of the selected model (*i.e.*, the selection model). We do so by using deleted residuals on the respondents' BMI values.

Let $(\boldsymbol{x}_{(ij)}, \boldsymbol{r}_{(ij)})$ denote the vector of all observations excluding the $(i\,j)^{\text{th}}$ observation (x_{ij}, r_{ij}). Then, the $(i\,j)^{\text{th}}$ deleted residual is given by

$$\mathrm{DRES}_{ij} = \{x_{ij} - \mathrm{E}(x_{ij} \mid \boldsymbol{x}_{(ij)}, \boldsymbol{r}_{(ij)})\} \big/ \mathrm{STD}(x_{ij} \mid \boldsymbol{x}_{(ij)}, \boldsymbol{r}_{(ij)}).$$

These values are obtained by performing a weighted importance sampling on the Metropolis-Hastings output. The posterior moments are obtained from

$$f(x_{ij} \mid \boldsymbol{x}_{(ij)}, \boldsymbol{r}_{(ij)}) = \int f(x_{ij} \mid \Omega)\pi(\Omega \mid \boldsymbol{x}_{(ij)}, \boldsymbol{r}_{(ij)}) d\Omega.$$

For the pattern mixture model

$$\begin{aligned} f(x_{ij} \mid \Omega) = {} & f(x_{ij} \mid r_{ij} = 0, \Omega) p(r_{ij} = 0 \mid \Omega) \\ & + f(x_{ij} \mid r_{ij} = 1, \Omega) p(r_{ij} = 1 \mid \Omega) \end{aligned}$$

and for the selection model

$$f(x_{ij} \mid \Omega) \sim \text{Normal}\{(\boldsymbol{z}'_{ij}\boldsymbol{\alpha}_1 + v_{0i}) + (\boldsymbol{z}'_{ij}\boldsymbol{\alpha}_2 + v_{1i})a_{ij}, \sigma_3^2\}.$$

We also considered using the conditional posterior ordinate (CPO) which is $f(x_{ij} \mid \boldsymbol{x}_{(ij)}, \boldsymbol{r}_{(ij)})$ evaluated at the observed x_{ij}. However, these CPO's lead to similar results for identifying extremes.

We drew box plots (not shown) of DRES versus the four levels of race-sex and the thirty five counties, and they showed that the selection model fits well. We drew box plots of DRES versus age and, interestingly, we found a pattern. Age class 2–4 seems to fit well; the predicted BMI values are somewhat high for age class 5–9; and age classes 10–14 and 15–19 have larger variability. We look at the box plots of DRES versus age even further by separating out the box plots for 18 (*i.e.*, 2–19 years old) individual ages (see Figure 1). Ages 11–19 fits well, but there is a problem with ages 2–10 (*i.e.*, a downward curvature in the medians). The other three models show similar patterns. A further refinement of the selection model in section 5 fixes this problem.

4. Estimation and Prediction

In this section we perform an analysis on the NHANES III BMI data for children and adolescents (*i.e.*, 2–19 years old). We use the selection model, and then as a means to study sensitivity, we compare prediction under the non-ignorable nonresponse selection model with that of the other three models.

4.1 Estimation

We have studied the relation between BMI and age using 95% credible intervals for the parameters in the selection model. First, the interaction of race and sex is not important, but as expected there is an important relation of BMI on age. BMI increases substantially with age (95% credible interval for α_{21} is (11.89, 13.67)). The rate of increase for white males is smaller (95% credible interval for α_{22} is $(-2.30, -0.19)$ and the 95% credible interval for α_{23} is $(-3.03, -0.64)$). Thus, while BMI increases with age, there is relatively less increase for white males. Apart from the parameter θ_1, which indicates strong nonignorability, the other parameters are essentially unimportant. For example, the 95% credible intervals for ρ_1 and ρ_2 are $(-0.53, 0.39)$ and $(-0.45, 0.45)$ respectively indicating that a simpler model can be used (*i.e.*, $\rho_1 = \rho_2 = 0$).

We take up the issue of ignorability further. We drew box plots (not shown) of the posterior densities of the β_{1i}, obtained from the iterates from the Metropolis-Hastings sampler, by county. All the box plots are above zero. This suggests that the nonresponse mechanism for each county is nonignorable. In addition, there are varying degrees of nonignorability. For example, several counties have the medians of the box plots near 1.5 while others have them near 2.

4.2 Prediction

It is desirable to predict the finite population mean BMI value and the proportion of respondents in the finite population. The sampled nonrespondents' BMI values are obtained through their conditional posterior densities included in the Metropolis-Hastings sampler. The non-sampled BMI values are to be predicted.

It is worthwhile noting that our models are applied to the logarithm of BMI with each individual having her/his covariates, and so the logarithm of each individual non-sampled value has to be predicted and then retransformed to the original scale. However, the computation is reduced considerably because age, race and sex for each nonsampled individual is not known, but the number of individuals in each age-race-sex domain is known in the U.S. population by county.

The distributions of the nonsampled individuals are

$$f(x_{ij}, r_{ij} | \boldsymbol{x}^{\text{obs}}, \boldsymbol{r}^{\text{obs}}) = \int f(x_{ij}, r_{ij} | \Omega) \pi(\Omega | \boldsymbol{x}^{\text{obs}}, \boldsymbol{r}^{\text{obs}}) d\Omega,$$

$i = 1, \ldots, \ell,\ j = n_i + 1, \ldots, N_i$. For the pattern mixture model we have

$$f(x_{ij}, r_{ij} | \Omega) = f(x_{ij} | r_{ij}, \Omega) p(r_{ij} | \Omega)$$

and for the selection model we have

$$f(x_{ij}, r_{ij} | \Omega) = p(r_{ij} | x_{ij}, \Omega) f(x_{ij} | \Omega),$$

where Ω denote the set of all parameters.

Therefore, if we take a sample of size M from the posterior distribution, $\{\Omega^{(h)} : h = 1, \ldots, M\}$, an estimator for $f(x_{ij}, r_{ij} \mid \boldsymbol{x}^{\text{obs}})$

$$f(x_{ij}, \hat{r}_{ij} | \boldsymbol{x}^{\text{obs}}) = M^{-1} \sum_{h=1}^{M} f(x_{ij}, r_{ij} | \Omega^{(h)}).$$

Thus, we can fill in the x_{ij} and r_{ij} for each $\Omega^{(h)}$ obtained from the MCMC algorithm from which we get M realizations $\overline{X}_i^{(h)}, P_i^{(h)}, h = 1, \ldots, M$. Inference can now be made about $\overline{X}_i$ in (1) and P_i in (2).

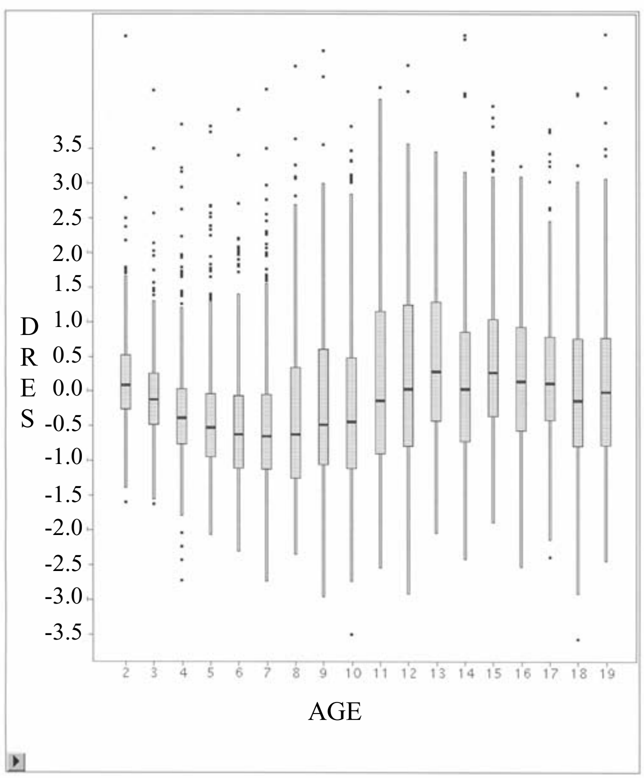


Figure 1. Box plots of the cross-validation residuals (DRES) by age for the selection model

We present 95% credible intervals for the finite population mean (FPM) BMI value and the finite population proportion (FPP) responding in order to judge sensitivity to the four models. Note that we provide these intervals for each domain: race by sex for each age class by county, and because they are very similar across domains we have presented in Table 2 the average of the end points of the credible intervals over county for black females only. The intervals for the FPM across the models are very similar. However, those for the FPP are very different. The intervals for the pattern mixture model and its ignorable version are similar except for age class 2–4. This is expected because these models express a linear regression of the logarithm of the odds of responding on age. The intervals for the FPP under the two pattern mixture models are essentially the same because they have the same relation with age, race, sex and their interaction. The intervals for the ignorable version of the selection model are all the same over age because in the response part of this model both age and BMI are ignored. We note that the intervals for the selection model have forms similar to the pattern mixture model and its ignorable version. As the intervals indicate, the FPM and FPP increase with age.

5. A Spline Regression Model

We now address the issue associated with the box plot in Figure 1. We have a further look at the observed data. A box plot of observed BMI values versus age shows that BMI is roughly constant for ages 2–8, then rises roughly linearly for ages 8–13, and finally rises very slowly for ages 14–19. This apparently important feature is not included in the four models. Thus, in this section we attempt to exploit this feature using a spline regression model.

We have used Part 1 of the selection model, and for Part 2 we use a join-point regression model. Generically, letting $c^+ = 0$ if $c \le 0$ and $c^+ = c$ if $c > 0$, we take

$$x_{ij} = \varphi_{0ij} + \varphi_{1ij}(a_{ij} - 8)^+ + \varphi_{2ij})a_{ij} - 13^+ + e_{ij} \qquad (9)$$

where in the spirit of our four models

$$\varphi_{kij} = \mathbf{z}_{ij}\, \boldsymbol{\alpha}_k + v_{ki}, \quad k = 0, 1, 2.$$

In (9) we have taken

$$e_{ij} \mid \sigma_3^2 \overset{\text{idd}}{\sim} \text{Normal}(0, \sigma_3^2)$$

and motivated by our earlier result (the v_{ki} are uncorrelated), rather than a trivariate normal density on $\boldsymbol{v}_i = (v_{1i}, v_{2i}, v_{3i})'$, we have taken

$$v_{ki} \mid \sigma_k^2 \overset{\text{idd}}{\sim} \text{Normal}(0, \sigma_k^2),\ k = 0, 1, 2.$$

The distribution assumptions on the hyper-parameters remain unchanged.

We have computed the deviance measure for the spline model; see Table 1 for the other four models. For this model $G \approx 129$ and $P \approx 107$ compared with $G \approx 118$ and $P \approx 135$ for the selection model. That is, $D_{100} \approx 236$ for the spline regression model and $D_{100} \approx 253$ for the selection model. Thus, the spline regression model shows an improvement over the original selection model.

In Figure 2 we present box plots of DRES versus age. This is a much improved plot over the one for the selection model (see Figure 1). Observe that the medians fluctuate about 0 with very little variation. The box plots for ages 2, 3, 4, 5, 6 and 7 are a little less variable than the others. We also fit the quadratic join-point model in which we replace (9) by

$$x_{ij} = \varphi_{0ij} + \varphi_{1ij}(a_{ij} - 8)^+ + \varphi_{2ij}\{(a_{ij} - 13)^+\}^2 + e_{ij}$$

with all other assumptions remaining unchanged. This model did not show any substantial improvement over the alternative model specified by (9), which we retain without further refinement.

Table 2
Comparison of the Four Models Based on the Average Over All Counties of the End Points of the 95% Credible Intervals for the Finite Population Mean BMI (FPM) and Proportion (FPP) Responding for Black Females

Model		age 2 – 4	5 – 9	10 – 14	15 – 19
SEI	FPM	(14.80, 16.07)	(17.09, 18.58)	(19.63, 21.61)	(22.40, 25.19)
	FPP	(0.73, 0.79)	(0.73, 0.79)	(0.73, 0.79)	(0.73, 0.79)
SE	FPM	(15.55, 16.21)	(17.49, 18.36)	(19.52, 20.92)	(21.74, 23.91)
	FPP	(0.66, 0.78)	(0.71, 0.81)	(0.75, 0.84)	(0.78, 0.87)
PMI	FPM	(14.75, 16.10)	(17.04, 18.59)	(19.59, 21.55)	(22.42, 25.09)
	FPP	(0.49, 0.70)	(0.72, 0.84)	(0.84, 0.94)	(0.90, 0.98)
PM	FPM	(14.96, 15.79)	(17.16, 18.38)	(19.61, 21.45)	(22.37, 25.07)
	FPP	(0.49, 0.70)	(0.73, 0.84)	(0.84, 0.94)	(0.90, 0.98)

Note: SEI is ignorable version of the selection model, PMI is ignorable version of the pattern mixture model, PM is pattern mixture model, and SE is selection model.

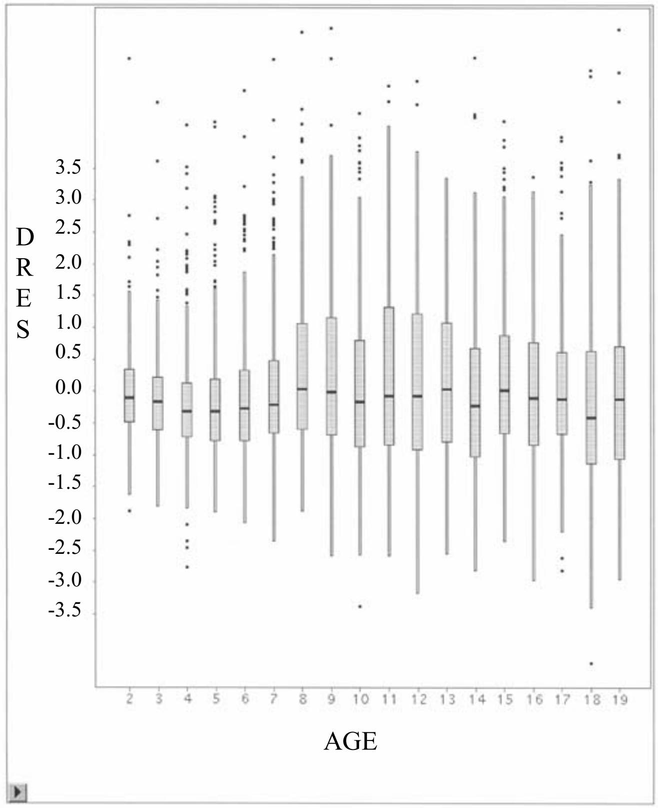


Figure 2. Box plots of the cross-validation residuals (DRES) by age for the spline regression model

In Table 3 we compare the FPM for the selection models (regression without splines and regression with splines). Again we average the end points of the 95% credible intervals over all counties. The intervals overlap suggesting similarity between the model without splines and the one with them. However, there are some exceptions. The largest difference between the intervals occur for individuals age 15–19 years old. In general, the spline model provides higher precision. For example, for age 10–19 the intervals for the spline model are contained by those for the model without the splines.

6. Conclusions

To analyze BMI data from NHANES III by age, race and sex within each county, (a) we have extended the normal-logistic regression model to a hierarchical Bayesian selection model, and (b) constructed a pattern mixture model and two ignorable nonresponse models to assess sensitivity to inference. A deviance measure shows that among the four models, the selection model is the best, and a cross-validation analysis shows that these models fit roughly equally well.

Table 3
Comparison of the Two Selection Models (Regression Without Splines and Regression with Splines) using the Average Over all Countries of the End Points of the 95% Credible Intervals for the Finite Population Mean BMI by Age, Race and Sex

		age			
R–S		2–4	5–9	10–14	15–19
BF	No Spline	(16.26, 16.92)	(16.44, 17.10)	(19.62, 21.41)	(21.35, 25.62)
	Spline	(15.65, 16.31)	(17.62, 18.41)	(19.70, 20.91)	(21.95, 23.82)
BM	No Spline	(16.10, 16.76)	(16.26, 16.92)	(18.83, 20.55)	(20.45, 24.53)
	Spline	(15.68, 16.32)	(17.32, 18.11)	(19.03, 20.21)	(20.84, 22.61)
OF	No Spline	(16.39, 17.00)	(16.56, 17.17)	(19.48, 21.19)	(21.16, 25.39)
	Spline	(16.01, 16.60)	(17.77, 18.54)	(19.62, 20.79)	(21.61, 23.38)
OM	No Spline	(16.53, 17.14)	(16.67, 17.29)	(19.22, 20.95)	(20.83, 24.98)
	Spline	(16.16, 16.74)	(17.74, 18.51)	(19.38, 20.55)	(21.13, 22.87)

Note: R–S is race-sex: BF is black female; BM is black male; OF is non-black female; and OM is non-black male.

Another contribution is the identification of a common deficiency in the selection model, the pattern mixture model and the two ignorable models. Based on the observed data, we have found that there is a dynamic relationship of BMI with age. Thus, we have further extended the selection model to include three linear splines. The cross validation analysis shows that there is an improvement over the selection model, and in fact, the deviance measure shows that the linear spline regression model is the best among the five models.

Our study on obesity is one of the key contributions in this work. The linear spline regression of BMI on age adjusting for race and sex, gives a better fit and improved precision than the selection model without splines. It is not easy to construct a model that is satisfactory for all aspects of the NHANES III data simultaneously. We have been able to do so for children and adolescents. BMI increases substantially with age; race and sex contributing negatively to this increase; there is relatively less increase for white males. In general, the effects of race and sex are relatively minor. There is some variation across the thirty five counties.

Appendix A
The Pattern Mixture Model

For Part 1 of the pattern mixture model the response depends on age, race and sex, and the interaction of race and sex through the logistic regression

$$r_{ij} \mid \beta_i \overset{\text{ind}}{\sim} \text{Bernoulli}$$
$$\left\{ e^{\beta_{0i}+\beta_{1i}a_{ij}+\beta_{2i}z_{ij1}+\beta_{3i}z_{ij2}+\beta_{4i}z_{ij3}} / (1+e^{\beta_{0i}+\beta_{1i}a_{ij}+\beta_{2i}z_{ij1}+\beta_{3i}z_{ij2}+\beta_{4i}z_{ij3}}) \right\}, \quad \text{(A.1)}$$

$i = 1, \ldots, l,\ j = 1, \ldots, N_i$. Now, letting, $\boldsymbol{\beta}_i = (\beta_{0i}, \beta_{1i}, \beta_{2i}, \beta_{3i}, \beta_{4i})'$, note that while the vector $\boldsymbol{\beta}_i$ has $p = 5$ components, the corresponding vector in (4) has two components. Analogous to (4) we take

$$\boldsymbol{\beta}_i \mid \boldsymbol{\theta}, \Delta \overset{\text{iid}}{\sim} \text{Normal}(\boldsymbol{\theta}, \Delta), \quad \text{(A.2)}$$

and for the prior distribution,

$$\boldsymbol{\theta} \sim \text{Normal}(\boldsymbol{\theta}^{(0)}, \Delta^{(0)})$$
$$\text{and} \quad \Delta^{-1} \sim \text{Wishart}\{(\nu^{(0)}\Lambda^{(0)})^{-1}, \nu^{(0)}\}, \nu^{(0)} > p, \quad \text{(A.3)}$$

where $\boldsymbol{\theta}^{(0)}, \Delta^{(0)}, \Lambda^{(0)}$ and $\nu^{(0)}$ are to be specified. Part 2 of this model for BMI incorporates a dependence on the response indicators, letting $w_{ij0} = 1, w_{ij1} = a_{ij}$,

$$x_{ij} = \sum_{t=0}^{1} (\mathbf{z}'_{ij}\boldsymbol{\alpha}_t + r_{ij}v_{ti})w_{ijt} + e_{ij},\ r_{ij} = 0, 1,$$
$$e_{ij} \mid \sigma_3^2 \overset{\text{iid}}{\sim} \text{Normal}(0, \sigma_3^2). \quad \text{(A.4)}$$

The distributions on the (v_{0i}, v_{1i}) are the same as in (7). The prior distributions are exactly those in Part 2 of the selection model (*i.e.*, see (6) and (7)).

We take $\nu^{(0)} = 2p$, a value that indicates near vagueness, maintains propriety and permits stability in computation. We show how to specify parameters like $\boldsymbol{\theta}^{(0)}, \Delta^{(0)}, \boldsymbol{\alpha}_t^{(0)}, \Delta_t^{(0)},\ t = 1, 2, 3, \Lambda^{(0)}$ in Appendix C.

Appendix B
Metropolis-Hastings Algorithm for Fitting the Selection Model

For the nonignorable nonresponse selection model the joint posterior density is

$$p(\boldsymbol{x}^{(s,nr)}, \boldsymbol{\sigma}^2, \boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{v}, \boldsymbol{\theta}, \rho_1, \rho_2 \mid \boldsymbol{x}^{(s,r)}) \propto$$
$$\prod_{i=1}^{l}\left\{\prod_{j=1}^{r_i} \frac{1}{\sigma_3} e^{-\frac{1}{2\sigma_3^2}(x_{ij}-\{\mathbf{z}'_{ij}(\boldsymbol{\alpha}_1+a_{ij}\boldsymbol{\alpha}_2)+v_{0i}+v_{1i}a_{ij}\})^2} \frac{e^{\beta_{0i}+\beta_{1i}x_{ij}}}{1+e^{\beta_{0i}+\beta_{1i}x_{ij}}}\right\}$$
$$\times\prod_{i=1}^{l}\left\{\prod_{j=r_i+1}^{n_i} \frac{1}{\sigma_3} e^{-\frac{1}{2\sigma_3^2}(x_{ij}-\{\mathbf{z}'_{ij}(\boldsymbol{\alpha}_1+a_{ij}\boldsymbol{\alpha}_2)+v_{0i}+v_{1i}a_{ij}\})^2} \frac{1}{1+e^{\beta_{0i}+\beta_{1i}x_{ij}}}\right\}$$
$$\times\left\{\begin{array}{l}\displaystyle\prod_{i=1}^{l} \frac{1}{\sigma_1\sigma_2\sqrt{1-\rho_1^2}} \\ e^{-\frac{1}{2(1-\rho_1^2)}\left[\left(\frac{\beta_{0i}-\theta_0}{\sigma_1}\right)^2-2\rho_1\left(\frac{\beta_{0i}-\theta_0}{\sigma_1}\right)\left(\frac{\beta_{1i}-\theta_1}{\sigma_2}\right)+\left(\frac{\beta_{1i}-\theta_1}{\sigma_1}\right)^2\right]}\end{array}\right\}$$
$$\times\left\{\prod_{i=1}^{l} \frac{1}{\sigma_4\sigma_5\sqrt{1-\rho_2^2}} e^{-\frac{1}{2(1-\rho_2^2)}\left[\left(\frac{v_{0i}}{\sigma_4}\right)^2-2\rho_2\left(\frac{v_{0i}}{\sigma_4}\right)\left(\frac{v_{1i}}{\sigma_5}\right)+\left(\frac{v_{1i}}{\sigma_5}\right)^2\right]}\right\}$$
$$\times\left\{\prod_{k=1}^{5}\left(\frac{1}{\sigma_k^2}\right)^{\frac{a}{2}+1} e^{-\frac{a}{2\sigma_k^2}}\right\}\left\{e^{-\frac{1}{2}(\boldsymbol{\theta}-\boldsymbol{\theta}^{(0)})'\Delta^{(0)-1}(\boldsymbol{\theta}-\boldsymbol{\theta}^{(0)})}\right\}$$
$$\times\left\{\prod_{k=1}^{2} e^{-\frac{1}{2}(\boldsymbol{\alpha}_k-\boldsymbol{\alpha}_k^{(0)})'\Delta_k^{(0)^{-1}}(\boldsymbol{\alpha}_k-\boldsymbol{\alpha}_k^{(0)})}\right\}. \quad \text{(B.1)}$$

Let Ω denote the set of parameters $\boldsymbol{\beta}, \boldsymbol{\theta}, \boldsymbol{v}, \boldsymbol{\alpha}, \sigma_3^2, \boldsymbol{\psi}_1, \boldsymbol{\psi}_2$ and $\boldsymbol{x}^{(s,nr)}$ where $\boldsymbol{\psi}_1 = (\sigma_1^2, \sigma_2^2, \rho_1)'$ and $\boldsymbol{\psi}_2 = (\sigma_4^2, \sigma_5^2, \rho_2)'$. Generically, let $\Omega_{\boldsymbol{a}}$ denote all parameters in Ω except $\boldsymbol{a}$; for example, $\Omega_{\boldsymbol{\beta}} = (\boldsymbol{\theta}, \boldsymbol{v}, \boldsymbol{\alpha}, \sigma_3^2, \boldsymbol{\psi}_1, \boldsymbol{\psi}_2, \boldsymbol{x}^{(s,nr)})$, so that the conditional posterior density (CPD) of $\boldsymbol{\beta}$ is denoted by $p(\boldsymbol{\beta} \mid \Omega_{\boldsymbol{\beta}}, \boldsymbol{x}^{(s,r)})$. To perform the Metropolis-Hastings algorithm, one needs the CPD for each parameter given the others and $\boldsymbol{x}^{(s,r)}$. Here we give a sketch of the algorithm.

The CPD for each of the parameters $\boldsymbol{\theta}, \boldsymbol{v}, \boldsymbol{\alpha}$ and σ_3^2 is easy to write down. But we need Metropolis steps for the CPD's of $\boldsymbol{\beta}, \boldsymbol{\psi}_1, \boldsymbol{\psi}_2$, and $\boldsymbol{x}^{(s,nr)}$.

Conditioning on $\Omega_{\boldsymbol{\beta}}$, the parameters $\boldsymbol{\beta}_1, \ldots, \boldsymbol{\beta}_l$, are independent with

$$p(\boldsymbol{\beta}_i \mid \boldsymbol{x}^{(s,r)}) \propto \prod_{j=1}^{n_i} \left\{ \frac{e^{(\beta_{0i}+\beta_{1i}x_{ij})r_{ij}}}{1+e^{(\beta_{0i}+\beta_{1i}x_{ij})}} \right\} \times e^{-\frac{1}{2}(\boldsymbol{\beta}_i-\boldsymbol{\theta})'\Delta_1^{-1}(\boldsymbol{\beta}_i-\boldsymbol{\theta})},$$

where

$$\Delta_1 = \begin{pmatrix} \sigma_1^2 & \rho_1\,\sigma_1\,\sigma_2 \\ \rho_1\,\sigma_1\,\sigma_2 & \sigma_2^2 \end{pmatrix}$$

and $x_{ij}, i=1,\ldots,l$ and $j=r_i+1,\ldots,n_i$ are to be predicted; see below. We use a technique based on logistic regression to obtain a multivariate Student's t proposal density in which tuning is obtained by varying its degree of freedom.

The method to draw from the CPD's of $\boldsymbol{\psi}_1 = (\sigma_1^2, \sigma_2^2, \rho_1)$ and $\boldsymbol{\psi}_2 = (\sigma_3^2, \sigma_4^2, \rho_2)$ is the same. The CPD of ψ_2 is

$$p(\boldsymbol{\psi}_2 \mid \Omega_{\boldsymbol{\psi}_2}, \boldsymbol{x}^{(s,r)}) \propto \left(\frac{1}{\sigma_4^2\,\sigma_5^2} \right)^{\frac{a+l}{2}+1} e^{-\frac{b}{2}\left(\frac{1}{\sigma_4^2}+\frac{1}{\sigma_5^2}\right)}$$

$$\times \frac{1}{(1-\rho_2^2)^{l/2}} e^{-\frac{1}{2(1-\rho_2^2)}\left\{\frac{1}{\sigma_4^2}\sum_{i=1}^{l} v_{0i}^2 - \frac{2\rho_2}{\sigma_4\sigma_5}\sum_{i=1}^{l} v_{0i}v_{1i} + \frac{1}{\sigma_5^2}\sum_{i=1}^{l} v_{1i}^2\right\}}.$$

We have used the Fisher's z transformation (see Ruben 1966) to obtain a proposal density associated with normal distribution for $\log\{\rho_2/(1-\rho_2)\}$ and gamma distributions for σ_4^2 and σ_5^2.

Finally, we consider the Metropolis step for drawing $\boldsymbol{x}^{(s,nr)} \mid \Omega_{\boldsymbol{x}^{(s,nr)}}, \boldsymbol{x}^{(s,r)}$. We note that in this CPD, $x_{ij}, i=1,\ldots,l,\ j=r_i+1,\ldots,n_i$, are independent with

$$p(x_{ij} \mid \Omega_{ij}, \boldsymbol{x}^{(s,r)}) \propto e^{-\frac{1}{2\sigma_3^2}[x_{ij}-\{\boldsymbol{z}_{ij}(\boldsymbol{\alpha}_1+a_{ij}\boldsymbol{\alpha}_2)+v_{0i}+v_{1i}a_{ij}\}]^2}$$
$$\left\{1+e^{\beta_{0i}+\beta_{1i}x_{ij}}\right\}^{-1}.$$

We have constructed a proposal density using least squares techniques. We note that the proposal density Normal $(\boldsymbol{z}_{ij}(\boldsymbol{\alpha}_1 + a_{ij}\boldsymbol{\alpha}_2) + v_{0i} + v_{1i}a_{ij}, \sigma_3^2)$ did not perform well (see Chib and Greenberg 1995).

Appendix C
Specification of Hyperparameters

We discuss how to specify the hyperparameters $(\boldsymbol{\theta}^{(0)}, \Delta^{(0)})$ and $(\boldsymbol{\alpha}_k^{(0)}, \Gamma_k^{(0)}), k=1,2$, associated with $\boldsymbol{\theta}$ and $\boldsymbol{\alpha}_k, k=1,2$ in the selection model.

First, consider $(\boldsymbol{\theta}^{(0)}, \Delta^{(0)})$. For $i=1,\ldots,l,\ j=1,\ldots,n_i$ fit the logistic regression model $r_{ij} \overset{\text{iid}}{\sim}$ Bernoulli $\{e^{\beta_{0i}+\beta_{1i}x_{ij}}(1+e^{\beta_{0i}+\beta_{1i}x_{ij}})^{-1}\}$, where x_{ij} are obtained by prediction (see Appendix A). Letting $\hat{\boldsymbol{\beta}}_i, i=1,\ldots,l$ denote the least squares estimators, we assume that $\hat{\boldsymbol{\beta}}_i \overset{\text{iid}}{\sim}$ Normal $(\boldsymbol{\theta}^{(0)}, \tilde{\Delta}^{(0)})$ to get $\boldsymbol{\theta}^{(0)} = 1/l\sum_{i=1}^{l}\hat{\boldsymbol{\beta}}_i$ and

$$\hat{\tilde{\Delta}}^{(0)} = \frac{1}{l-1}\sum_{i=1}^{l}(\hat{\boldsymbol{\beta}}_i - \boldsymbol{\theta}_{(0)})(\hat{\boldsymbol{\beta}}_i - \boldsymbol{\theta}_{(0)}) \qquad \text{(C.1)}$$

and we set $\Delta^{(0)} = \kappa_1\hat{\tilde{\Delta}}^{(0)}$, where κ_1 is to be selected.

Next, we consider how to specify $(\boldsymbol{\alpha}_k^{(0)}, \Gamma_k^{(0)}), k=1,2$. We fit $x_{ij} = \boldsymbol{z}'_{ij}(\boldsymbol{\alpha}_1 + \boldsymbol{\alpha}_2 a_{ij}) + e_{ij}$, where a_{ij} is the age of the j^{th} individual in the i^{th} county, $i=1,\ldots,l,\ j=1,\ldots,n_i$ to get least squares estimators, $\hat{\boldsymbol{\alpha}} = (\hat{\boldsymbol{\alpha}}_1, \hat{\boldsymbol{\alpha}}_2)$ and its covariance matrix $\hat{\Gamma}^{(0)}$. We set $\boldsymbol{\alpha}_k^{(0)} = \hat{\boldsymbol{\alpha}}_k$, and $\Gamma_k^{(0)} = \kappa_2\hat{\Gamma}_k^{(0)}$, where $\hat{\Gamma}_k^{(0)}, k=1,2$ is the corresponding block matrix of $\hat{\Gamma}^{(0)}, k=1,2$ and κ_2 is to be specified.

We have experimented with κ_1 in (C.1). We used $\kappa_1 = 100$ to provide a proper diffuse prior; a value of $\kappa_1 = 1{,}000$ did not change our predictions. Similarly, we used $\kappa_2 = 100$.

References

Chib, S., and Greenberg, E. (1995). Understanding the metropolis-hastings algorithm. *The American Statistician*, 49, 327-335.

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

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

Daniels, M.J. (1999). A prior for the variance in hierarchical models. *The Canadian Journal of Statistics*, 27, 569-580.

David, M., Little, R.J.A., Samuel, M.E. and Triest, R.K. (1986). Alternative methods for CPS income imputation. *Journal of the American Statistical Association*, 81, 29-41.

Dietz, W.H. (1998). Health consequences of obesity in youth: Childhood predictors of adult disease. *Pediatrics*, 101, 518-525.

Fagot-Campagna, A. (2000). Emergence of type 2 diabetes mellitus in children: Epidemiological evidence. *Journal of Pediatric Endocrinology Metabolism*, 13, 1395-1405.

Gelfand, A., and Ghosh, S. (1998). Model choice: A minimum posterior predictive approach. *Biometrika*, 85, 1-11.

Gelfand, A., Sahu, S. and Carlin, B. (1995). Efficient parametrisations for normal linear mixed models. *Biometrika*, 82, 479-488.

Gelman, A., Roberts, G.O. and Gilks, W.R. (1996). Efficient Metropolis jumping rules. In *Bayesian Statistics* (Eds.J.M. Bernardo, J.O. Berger, A.P. Dawid and A.F.M. Smith, Oxford, U.K.: Oxford University Press, 599-607.

Greenlees, J.S., Reece, W.S. and Zieschang, K.D. (1982). Imputation of missing values when the probability of response depends on the variable being imputed. *Journal of the American Statistical Association*, 77, 251-261.

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.

Laud, P., and Ibrahim, J. (1995). Predictive model selection. *Journal of the Royal Statistical Society, Series B*, 57, 247-262.

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

Mohadjar, L., Bell, B. and Waksberg, J. (1994). National health and Nutrition Examination Survey III-Accounting for item nonresponse bias. Internal Report, National Center for Health Statistics.

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

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

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.

Natarajan, R., and Kass, R.E. (2000). Reference Bayesian methods for generalized linear models. *Journal of the American Statistical Association*, 95, 227-237.

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.

Ogden, C.L., Flegal, K.M., Carroll, M.D. and Johnson, C.L. (2002). Prevalence and trends in overweight among us children and adolescents, 1999-2000. *Journal of the American Medical Association*, 288, 1728-1732.

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

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

Ruben, H. (1966). Some new results on the distribution of the sample correlation coefficient. *Journal of the Royal Statistical society, Series B*, 28, 513-525.

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

Rubin, D.B. (1977). Formalizing subjective notions about the effect of nonrespondents in sample surveys. *Journal of the American Statistical Association*, 72, 538-543.

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.

Squires, S. (2001). National plan urges to combat obesity: Weight-related illnesses kill 300,000 Americans annually, Surgeon General says. *The Washington Post*, December 14, 2001.

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.

Who Consultation on Obesity (2000). Obesity: Preventing and managing the global epidemic. *WHO Technical Report Series 894*, Geneva, Switzerland: World Health Organization.

Wright, C.M., Parker, L., Lamont, D. and Craft, A.W. (2001). Implications of childhood obesity for adult health: Findings from thousand families cohort study. *British Medical Journal*, 323, 1280-1284.