A Bayesian Assessment of Obesity Among Children and Adolescents from Small Domains Under Nonignorable Nonresponse

Balgobin Nandram[1], Jai Won Choi[2]

Department of Mathematical Sciences, Worcester Polytechnic Institute[1]

National Center for Health Statistics[2]

Abstract

We analyze body mass index (BMI) data from the Third National Health and Nutrition Examination Survey (NHANES III). However, because there are no BMI values for a considerable number of the children and adolescents we study, there can be serious nonresponse bias in inference. We construct two hierarchical Bayesian models, a selection model and a pattern mixture model including a spline regression, to analyze poststratified BMI data by age, race and sex within county accounting for these nonrespondents. We predict the finite population percentile BMI for small domains formed by age, race and sex in each of thirty five large counties. We use Markov chain Monte Carlo methods to fit the models, and a deviance measure to show that the selection model is better than the pattern mixture model under an appropriate transformation of the BMI values. We also study the goodness of fit of the models using a cross-validation and summarize the posterior predictive distribution of the finite population percentile of BMI by age, race and sex within county.

Keywords: Deleted residual; Deviance; Finite population percentiles; Logistic regression; Metropolis-Hastings sampler; Spline regression; Transformation.

1   Introduction

One of the variables in NHANES III is the body mass index (BMI, body weight in kilograms divided by [(height in meters)^2]) which is currently used as a measure to diagnose overweight and obesity in children and adolescents. Youths with BMIs at least the 95th percentile for age and sex, are considered obese and referred for in depth medical follow-up to determine underlying diagnoses. Those with BMIs at least the 85th percentile but less than the 95th percentile are be considered at risk of overweight, and should be referred to a second-level screen.

The NHANES III data are biased because there are many nonrespondents among children and adolescents, and the main issue we address here is that nonresponse should not be ignored because respondents may differ from nonrespondents. The purpose of this work is to predict the percentile BMI for the finite population of [1]

children and adolescents, poststratified 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 two hierarchical Bayesian models to accommodate the nonresponse mechanism. Nandram and Choi (2005) develop hierarchical Bayesian models for obesity. The main contribution in their paper is Bayesian predictive inference of the finite population mean using a spline regression model in which the logarithm of the BMI values are transformed. Our contributions here are different. First, we make inference about finite population percentiles. Second, we show that the logarithmic transformation is the best in a selected set within the Box-Cox family. Third, we demonstrate the effects of clustering and survey weights.

Nonresponse models can be classified very broadly (e.g., see Little and Rubin 1987). Let $[x]$ and $[r]$ denote respectively the density function of the response variable x, and the response indicator r, with obvious notations for the joint and conditional densities. Then the selection model specifies that $[x, r] = [r \mid x][x]$ and the pattern mixture model specifies $[x, r] = [x \mid r][r]$. While the two models have the same joint density, in practice the components $[r \mid x]$ and $[x]$ for the selection model, and $[x \mid r]$ and $[r]$ for the pattern mixture model are specified. Thus, it is not surprising that these models are different. If $[x, r] = [x][r]$ and the parameters of the density function $[x]$ are unrelated to those of the mass function $[r]$, the model is said to be ignorable; otherwise it is nonignorable. We use both selection and pattern mixture models in the hierarchical Bayesian framework for our nonignorable nonresponse problem. Greenlees et. al (1982) developed a normal-logistic regression model, a nonignorable nonresponse model within the selection approach, we extend this model to accommodate clustering for the NHANES III data, and incorporate a regression with linear splines to accommodate a dynamic relation between BMI and age, race, sex, and race sex interaction.

The purpose of this paper is primarily to report on an innovative Bayesian study of obesity and overweight in children and adolescents using continuous BMI data obtained for thirty five counties in NHANES III, accounting for the nonresponse. The rest of the paper is organized as follows. In Section 2, we discuss the hierarchical Bayesian methodology for nonignorable nonresponse through the selection and the pattern mixture approaches. We also

---

[1]The opinions expressed in this paper are those of the authors and do not necessarily represent the views of the National Center for Health Statistics

describe model fitting, selection and assessment including predictive deviance and a cross-validation. In Section 3 we compare the selection and the pattern mixture models, and we describe the data analysis and a simulation study. Finally, Section 4 has a brief discussion about how to include clustering among the households and survey weights in our selection model (i.e., the selected model).

2 Hierarchical Bayesian Methodology

In this section we describe two Bayesian models for nonignorable nonresponse, model selection and assessment for the selected 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 ith county. Let s denote the set of sampled units and ns the set of nonsampled units. Let r_{ij} for $i = 1, 2, ..., \ell$ and $j = 1, 2, ..., N_i$ be the response indicator for the jth individual within the ith county in the population. Also, let x_{ij} denote the BMI value, possibly transformed (e.g., the logarithmic transformation). 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 ith 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 i^{th} county.

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 the $100\eta, 0 < \eta < 1$, percentile of the finite population of BMI values for each age-race-sex domain within the ith county. For example, for the ith county, let $\mathbf{x}_i = (\mathbf{x}_i^{(s,r)}, \mathbf{x}_i^{(s,nr)}, \mathbf{x}_i^{(ns)})'$, where $\mathbf{x}_i^{(s,r)}$ is observed BMI values of the sampled respondents, and both $\mathbf{x}_i^{(s,nr)}$, the BMI values of the sampled nonrespondents, and $\mathbf{x}_i^{(ns)}$, the nonsampled BMI, are not observed. Then, the $[\eta N_i]^{th}$ percentile ([·] is the nearest integer to ηN_i) of the ith county is the $[\eta N_i]^{th}$ order statistic among the N_i components of $\mathbf{x}_i$. Because only $\mathbf{x_i}^{(s,r)}$ is observed, we develop a Bayesian selection and a Bayesian pattern mixture model to predict the finite population percentile BMI for each domain.

2.1 Competing Models

We describe the selection and the pattern mixture model. To accommodate the BMI values and the nonresponse indicators, each model has two parts which are combined probabilistically.

The single most important predictor of BMI is age, with race and sex playing a relatively minor role, and there is a need to understand the relationship between BMI and age, race and sex. For $i = 1, \ldots, \ell$, $j =$ $1, \ldots, N_i$, 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 $\mathbf{z}'_{ij} = (z_{ij0}, z_{ij1}, z_{ij2}, z_{ij3})$. Also, let a_{ij} denote the age of the jth individual within the ith county. Generically, letting $c^+ = 0$ if $c \le 0$ and $c^+ = c$ if $c > 0$, $w_{ij1} = 1$, $w_{ij2} = (a_{ij} - 8)^+$, $w_{ij3} = (a_{ij} - 13)^+$, for a spline regression of BMI on age adjusting for race and sex, we take

$$x_{ij} = \sum_{t=1}^{3} (\mathbf{z}'_{ij}\alpha_t + \nu_{ti})w_{ijt} + e_{ij}, \ \ e_{ij} \mid \sigma_3^2 \overset{iid}{\sim} \mathrm{N}(0, \ \sigma_3^2), \ (1)$$

where N stands for normal and the ν_{ti} are random effects, and the regression coefficients, $\alpha_1' = (\alpha_{11}, \alpha_{12}, \alpha_{13}, \alpha_{14})$, $\alpha_2' = (\alpha_{21}, \alpha_{22}, \alpha_{23}, \alpha_{24})$, and $\alpha_3' = (\alpha_{31}, \alpha_{32}, \alpha_{33}, \alpha_{34})$, are of interest.

To account for variation across counties, note that in (1) for simplicity we only permit the α_{t1} to be random coefficients (i.e., the intercept in $\mathbf{z}'_{ij}\alpha_t + \nu_{ti}$ is $\alpha_{t1} + \nu_{ti}$). A random coefficients structure is included in (2) below.

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

$$r_{ij} \mid x_{ij}, \beta_i \overset{ind}{\sim} \mathrm{Bernoulli}\left\{ e^{\beta_{0i}+\beta_{1i}x_{ij}} / (1 + e^{\beta_{0i}+\beta_{1i}x_{ij}}) \right\},$$

$$\begin{pmatrix} \beta_{0i} \\ \beta_{1i} \end{pmatrix} \mid \Omega \overset{iid}{\sim} \mathrm{N} \left\{ \begin{pmatrix} \theta_0 \\ \theta_1 \end{pmatrix}, \begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix} \right\} \ (2)$$

where $\Omega = (\theta_0, \theta_1, \sigma_1^2, \sigma_2^2, \rho)$,

$$\theta \sim N(\theta^{(0)}, \Delta^{(0)}), \sigma_1^{-2}, \sigma_2^{-2} \overset{iid}{\sim} \mathrm{G}\left(a/2, \ a/2\right)$$

$$\text{and } \rho \sim \mathrm{U}(-1, 1),$$

G stands for Gamma, U stands for Uniform, and a, $\theta^{(0)}$ and $\Delta^{(0)}$ are to be specified. In Part 2 of the selection model, we start with (1), and we assume that

$$\nu_{ti} \mid \sigma_{t+3}^2 \overset{iid}{\sim} \mathrm{Normal}(0, \ \sigma_{t+3}^2), \ t = 1, 2, 3. \qquad (3)$$

For this part of the model, we use the prior distributions, for t=1,2,3,

$$\alpha_t \overset{iid}{\sim} \mathrm{N}(\alpha_t^{(0)}, \Delta_t^{(0)}) \text{ and } \sigma_{t+3}^{-2} \overset{iid}{\sim} \mathrm{G}\left(a/2, \ a/2\right), \qquad (4)$$

where a, $\alpha_t^{(0)}$ and $\Delta_t^{(0)}$, are to be specified.

Next, we describe the **pattern mixture model**. For Part 1 of this model the response depends on age, race and sex, and the interaction of race and sex through the logistic regression, letting $A_{ij} = (\beta_{0i} + \beta_{1i}a_{ij} + \beta_{2i}z_{ij1} + \beta_{3i}z_{ij2} + \beta_{4i}z_{ij3}$,

$$r_{ij} \mid \beta_i \overset{ind}{\sim} \mathrm{Bernoulli}\left\{ e^{A_{ij}} / (1 + e^{A_{ij}}) \right\}.$$

Now, letting $\beta_i = (\beta_{0i}, \beta_{1i}, \beta_{2i}, \beta_{3i}, \beta_{4i})'$, note that while the vector β_i has $p = 5$ components, the corresponding vector in (2) has two components. Analogous to (2), we take

$$\beta_i \mid \theta, \Delta \overset{iid}{\sim} \mathrm{Normal}(\theta, \Delta),$$

and for the prior distribution,

$$\theta \sim \text{Normal}(\theta^{(0)}, \ \Delta^{(0)}) \ \text{and}$$

$$\Delta^{-1} \sim \text{Wishart}\{(\nu^{(0)}\Lambda^{(0)})^{-1}, \ \nu^{(0)}\}, \ \nu^{(0)} > p,$$

where $\theta^{(0)}$, $\Delta^{(0)}$, $\Lambda^{(0)}$ and $\nu^{(0)}$ are to be specified. Part 2 of this model for BMI extends (1) to incorporate a dependence on the response indicators,

$$x_{ij} = \sum_{t=1}^{3}(\mathbf{z}'_{ij}\alpha_t + r_{ij}\nu_{ti})w_{ijt} + e_{ij}, r_{ij} = 0, 1,$$

$$e_{ij} \mid \sigma_3^2 \overset{iid}{\sim} \text{Normal}(0, \ \sigma_3^2). \tag{5}$$

This pattern mixture model is in the spirit of Rubin (1977). Note that while in the pattern mixture model in (5) there are two specifications/patterns for x_{ij} (i.e., $r_{ij} = 0$ and $r_{ij} = 1$) in the selection model in (1) there is a single specification. The prior distributions are exactly those in Part II of the selection model (i.e., see (3) and (4)).

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 $\theta^{(0)}$, $\Delta^{(0)}$, $\alpha_t^{(0)}$, $\Delta_t^{(0)}$, $t = 1, 2, 3$, $\Lambda^{(0)}$ in another place. 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 makes little difference in our analysis; see Nandram and Choi (2005) for further details.

2.2 Model Fitting

In this section we describe how to use the Metropolis-Hastings sampler to fit the models. Also a deviance measure is used to select one of the two models. Then, a cross-validation analysis is used to assess the goodness of fit of the selected model, and because the same general principle applies to both 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 BMI values of the nonrespondents in the sample (i.e., x_{ij} for $j = r_i + 1, \ldots, n_i$) are unknown and they are treated as parameters (i.e., they are latent variables). We denote these latent variables by $\mathbf{x}^{(s,nr)}$ and the observed data are denoted by $\mathbf{x}^{(obs)} \equiv \mathbf{x}^{(s,r)}$ (discussed earlier). For the selection model, using Bayes' theorem and letting Ω denote the set of parameters $\beta, \theta, \nu, \alpha, \sigma_3^2, \psi_1, \psi_2$ and $\mathbf{x}^{(s,nr)}$ where $\psi_1 = (\sigma_1^2, \sigma_2^2, \rho)'$ and $\psi_2 = (\sigma_4^2, \sigma_5^2, \sigma_6^2)'$, the joint posterior density is, letting $Q = (\frac{\beta_{0i}-\theta_0}{\sigma_1})^2 - 2\rho(\frac{\beta_{0i}-\theta_0}{\sigma_1})(\frac{\beta_{1i}-\theta_1}{\sigma_2}) + (\frac{\beta_{1i}-\theta_1}{\sigma_2})^2$,

$$p(\Omega \mid \mathbf{x}^{(s,r)}, \mathbf{r}) \propto$$

$$\prod_{i=1}^{\ell}\prod_{j=1}^{n_i}\left\{\sigma_3^{-1}e^{-\frac{1}{2\sigma_3^2}\{x_{ij}-\sum_{t=1}^{3}(\mathbf{z}'_{ij}\alpha_t + \nu_{it})w_{ijt}\}^2}\right.$$

$$\left. \times e^{r_{ij}(\beta_{0i}+\beta_{1i}x_{ij})}/(1+e^{\beta_{0i}+\beta_{1i}x_{ij}})\right\}$$

$$\times \prod_{i=1}^{\ell}\{\sigma_1^{-1}\sigma_2^{-1}(1-\rho^2)^{-1/2}e^{-\frac{1}{2(1-\rho^2)}Q}\}$$

$$\times e^{-\frac{1}{2}(\theta-\theta^{(0)})'\Delta^{(0)-1}(\theta-\theta^{(0)})}\prod_{t=1}^{3}\prod_{i=1}^{\ell}\{\sigma_{t+3}^{-1}e^{-\frac{\nu_{it}^2}{2\sigma_{t+3}^2}}\}$$

$$\times \prod_{t=1}^{3}\{e^{-\frac{1}{2}(\alpha_t-\alpha_t^{(0)})'\Delta_t^{(0)-1}(\alpha_t-\alpha_t^{(0)})}\}$$

$$\times \prod_{k=1}^{6}\{(\sigma_k^{-2})^{\frac{a}{2}+1}e^{-\frac{a}{2\sigma_k^2}}\}. \tag{6}$$

The posterior density in (6) is complex so that we use Markov chain Monte Carlo (MCMC) methods to draw samples from it; see Nandram and Choi (2005) for details. 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). We obtained a sample of 1000 iterates which we used for inference and model checking. 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 mostly in (.25, .50) (see Gelman et al. 1996).

2.3 Model Selection and Assessment

We used the minimum posterior predictive loss approach (Gelfand and Ghosh 1998) to select the better model. Under squared error loss the minimum posterior predictive loss is

$$D_k = P + \frac{k}{k+1}G \ \text{with} \ P = \sum_{ij}Var(x_{ij}^{(pre)} \mid \mathbf{x}^{(obs)})$$

$$\text{and} \ G = \sum_{ij}\left\{E(x_{ij}^{(pre)} \mid \mathbf{x}^{(obs)}) - x_{ij}^{(obs)}\right\}^2,$$

where

$$f(x_{ij}^{(pre)} \mid \mathbf{x}^{(obs)}) = \int f(x_{ij}^{(pre)} \mid \Omega)\pi(\Omega \mid \mathbf{x}^{(obs)})d\Omega$$

and $x_{ij}^{(pre)}$ are the predicted values and Ω is the set of all parameters. This measure extends one obtained earlier (Laud and Ibrahim 1995), and we have taken $k = 100$ to match this earlier version.

In Table 1 we present the deviance measure (D_{100}) and its associated components, goodness of fit (G) and

Table 1: Comparison of the selection and pattern mixture models using the deviance measure by transformation. Sel=selection, Pat=pattern mixture, Ide-identity, Log=Logarithm, SqR=Square root, CuR=Cube root, Squ=square, Cub=Cube, Type=Transformation type

Type	Selection			Pattern Mixture		
	P	G	D	P	G	D
Ide	4023	3475	7498	6197	3115	9312
Log	3827	3275	7102	5877	2949	8826
SqR	3927	3365	7292	6008	3017	9026
CuR	3875	3325	7200	5959	2991	8949
Squ	4271	3772	8043	6718	3390	10108
Cub	4536	4123	8659	7385	3730	11115

NOTE:Identity transformation uses the observed data directly (i.e., no transformation). $D_{100} = G + (\frac{100}{100+1})P$ where $G'd$ for goodness of fit, Pen for penalty and $Dev \equiv D_{100}$ for deviance. We have selected the selection model with the logarithm transformation (see Section 2.3).

the penalty (P) for the selection and the pattern mixture models and six transformations (including the identity). For each transformation the selection model fits better (i.e., smaller G); for example, under the logarithmic transformation the selection model fits better (true for all transformations): compare D_{100} (7102 vs. 8826), P (3827 vs. 5877) and G (3275 vs. 2949) which, however, is a bit bigger for the selection model. We select the best transformation within the set of six transformations (see Table 1) for the selection model using a probabilistic argument. We obtain the posterior probability for each transformation assuming a priori that the six transformations are equally likely. The logarithm transformation gives the highest posterior probability, followed by the square root transformation. Also, we have found that the 95% credible intervals for the percentiles of population BMI values under the logarithmic and square root transformations are very similar.

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

$$\text{DRES}_{ij} = \frac{x_{ij} - \text{E}\left(x_{ij} \mid \mathbf{x}_{(ij)}, \mathbf{r}_{(ij)}\right)}{\text{STD}\left(x_{ij} \mid \mathbf{x}_{(ij)}, \mathbf{r}_{(ij)}\right)},$$

where we let $PRED = \text{E}\left(x_{ij} \mid \mathbf{x}_{(ij)}, \mathbf{r}_{(ij)}\right)$. These $DRES$ values are obtained by performing a weighted importance sampling on the output of the Metropolis-Hastings sampler. The posterior moments are obtained from

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

where $x_{ij} \mid \Omega \sim \text{Normal}\{\sum_{t=1}^{3}(\mathbf{z}'_{ij}\alpha_t + \nu_{ti})w_{ijt}, \ \sigma_3^2\}$ for the selection model (see (1)), and $f(x_{ij} \mid \Omega) = \sum_{r_{ij}=0}^{1} f(x_{ij} \mid r_{ij}, \Omega)p(r_{ij} \mid \Omega)$ for the pattern mixture model (see (5)).

We drew box plots (not shown) of DRES versus the four levels of race-sex and the thirty five counties, and they showed that both models fit well. Box plots of DRES versus the eighteen values (2-19 years) of age also show acceptable fit. We also plotted DRES versus PRED; both models show good ft, but the selection model appears to fit slightly better. Thus, based on the deviance and the box plots, we have selected the selection model.

3 Analysis, Prediction and Simulation

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 inference about the regression parameters from the selection and the pattern mixture models. We can provide 95% credible intervals for the finite population percentile for poststratified domains, but we report a selected set. We also perform a simulation study to assess the statistical features of the selection model.

3.1 Analysis

We have compared inference about the regression coefficients of BMI on age for both the selection model and the pattern mixture model. We present the posterior mean (PM) and 95% credible intervals (CI) in Table 2. Note that the CI for both models are small, showing that the estimates can be reproduced reasonably well. Also, note that some CIs from the selection model are slightly larger than those from the pattern mixture model, but the point estimates PMs are somewhatsimilar.

There are some differences between inference in the selection and mixture models. First, while for the selection model the interval for α_{12} (α_{22} as well) is to the left of 0, for the pattern mixture model the interval contains 0. The reverse is true for α_{33}. One important common feature is that both models show a substantial increase of BMI with age for children 8-13 years old and have the

Table 2: Posterior mean-PM, and 95% credible interval-CI for regression coefficients in spline part of 4 models.

coef.	PM	CI	PM	CI
		(a) Selection		(b) Pat. Mixture
α_{11}	-.767	(-.834,-.699)	-.524	(-.584,-.466)
α_{12}	-.084	(-.155,-.011)	.047	(-.024, .116)
α_{13}	-.025	(-.108, .060)	-.026	(-.108, .052)
α_{14}	.064	(-.044, .163)	.074	(-.019, .176)
α_{21}	.345	(.312, .381)	.311	(.279, .340)
α_{22}	-.021	(-.064,-.022)	-.032	(-.071, .007)
α_{23}	-.055	(-.100,-.011)	-.065	(-.107,-.024)
α_{24}	.011	(-.049, .069)	.027	(-.024, .084)
α_{31}	-.290	(-.367,-.223)	-.269	(-.337,-.199)
α_{32}	.049	(-.044, .145)	.045	(-.038, .127)
α_{33}	.077	(-.020, .173)	.098	(.008, .182)
α_{34}	-.006	(-.130, .122)	-.046	(-.168, .071)
		(c) InterClass Corr.		(d) Survey Weights
α_{11}	-.807	(-.873,-.737)	-.516	(-.911,-.091)
α_{12}	-.081	(-.161,-.001)	.014	(-.490, .472)
α_{13}	-.016	(-.099, .063)	-.035	(-.559, .463)
α_{14}	.066	(-.039, .168)	.050	(-.606, .681)
α_{21}	.339	(.308, .370)	.188	(-.054, .422)
α_{22}	-.029	(-.068, .011)	.001	(-.296, .302)
α_{23}	-.070	(-.110,-.031)	-.017	(-.344, .299)
α_{24}	.024	(-.032, .076)	.010	(-.384, .436)
α_{31}	-.297	(.365,-.231)	-.033	(-.598, .514)
α_{32}	.071	(-.019, .151)	-.005	(-.717, .691)
α_{33}	.113	(.025, .203)	-.001	(-.782, .748)
α_{34}	-.035	(-.156, .083)	-.005	(-.980, .943)

The components of α_{jk} correspond to $j = 1, 2, 3$ for ω_{ij1}, ω_{ij2} and ω_{ij3} and to $k = 1, 2, 3, 4$ for an intercept, race, sex, and the interaction of race and sex. In the spline model of ith county, we have used $(y_{ij} - \bar{y}_i)/s_i$ for the dependent variable (BMI) where $y_{ij} = \log(x_{ij})$, $\bar{y}_i = \sum_{j=1}^{r_i} y_{ij}/r_i$, and $s_i^2 = \sum_{j=1}^{r_i}(y_{ij} - \bar{y}_i)^2/(r_i - 1)$. It is not possible to provide inference on the original scale for α_1, α_2 and α_3, so the estimates are reported for the centered logarithm of the BMI values.

intervals for α_{31} to the left of 0, and the intervals for α_{14}, α_{24} and α_{34} contain 0. Thus, using the selection model there is substantial increase of BMI with age for children 8-13 years old, with a smaller increase for white males, and this increase is reduced significantly for adolescents 13-19 years old, without substantial adjustment for race and sex.

We take up the issue of ignorability further under the selection model. 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, and there are varying degrees of nonignorability; several counties have the medians of the box plots near 1.5 while others have them near 2.

3.2 Prediction

We obtain a summary (e.g., an upper percentile) of BMI values of the finite population (i.e., age-race-sex domain or a county) by performing a Bayesian predictive inference. The BMI values of the sampled nonrespondents are obtained through their conditional posterior densities included in the Metropolis-Hastings sampler. Thus, we describe how to predict the nonsampled BMI values.

Suppose we want to predict $\mathbf{x}^{(ns)} = (x_{ij}, \ j = r_i + 1, \ldots, N_i)$. Then,

$$f(\mathbf{x}^{(ns)} \mid \mathbf{x}^{(obs)}) =$$

$$\int \{ \prod_{j=r_i+1}^{N_i} f(x_{ij} \mid \Omega) \} \pi(\Omega \mid \mathbf{x}^{(obs)}) d\Omega,$$

where $x_{ij} \mid \Omega \sim \text{Normal}\{\sum_{t=1}^{3}(\mathbf{z}'_{ij}\alpha_t + \nu_{ti})w_{ijt}, \ \sigma_3^2\}$ for the selection model (see (1)), and $f(x_{ij} \mid \Omega) = \sum_{r_{ij}=0}^{1} f(x_{ij} \mid r_{ij}, \Omega) p(r_{ij} \mid \Omega)$ for the pattern mixture model (see (5)). Thus, it is straight forward to predict x_{ij} by taking a sample of size M from the posterior distribution, $\{\Omega^{(h)} : h = 1, \ldots, M\}$. (Once the transformed BMI values are predicted, thay can be easily retransformed to the original scale.)

Letting $\mathbf{x}_i^{(h)}$ denote the vector of all N_i iterated values, we order these components to obtain the $[100\alpha N_i]^{th}$ value, $x_i^{(h)}$, $h = 1, \ldots, M$, a "random sample" from the posterior density of the $[100\alpha N_i]^{th}$ percentile. In a similar manner a sample is obtained from the posterior distribution of any small area (e.g., white males 15-19 years old). We present 95% credible intervals for the finite population percentile (i.e., 85th, 95th) BMI values. These are obtained using the logarithm transformation (retransformed) and both the selection and pattern mixture models for each age-race-sex domain and each county. The differences among the race-sex groups are small (e.g., about one BMI points). For example, for county 11 95% credible intervals for the 95th percentile for age 5-9 corresponding to white males, black males, white females, black females are (19.2, 20.1),(20.6, 22.2),

(19.2, 20.1),(20.1,21.5). However, there are considerable differences among the counties, and among the thirty-five counties, there are smaller clusters of counties, with similarity within clusters. For example, in Table 3 we haved selected three counties in which counties 23 and 27 are similar, but county 11 is different from countis 23 and 27.

3.3 Simulation Study

We assess the predictive performance of the selection model, and therefore we simulate the entire population of values r_{ij} and x_{ij} from the selection model.

In the simulation we fix the parameters α_t, $t = 1, 2, 3$, σ_k^2, $k = 1, \ldots, 5$, ρ, θ_r, $r = 0, 1$ at their posterior means under the observed data. We kept the identifiers for race and sex indicators, n_i and N_i the same as in the original data. This design allows us to produce simulated data that are similar to the original, in particular, with the same missing data structure.

Then, we generate the r_{ij} and the x_{ij} according to the following scheme. First, we generate the β_i and the ν_i from

$$\begin{pmatrix} \beta_{0i} \\ \beta_{1i} \end{pmatrix} \overset{iid}{\sim} \text{N} \left\{ \begin{pmatrix} \theta_0 \\ \theta_1 \end{pmatrix}, \begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix} \right\}$$

and $\nu_{it} \overset{iid}{\sim} \text{Normal}(0, \ \sigma_{3+t}^2)$, $t = 1, 2, 3$,

$i = 1, \ldots, \ell$. Then we generate the r_{ij} and x_{ij} by drawing from

$$r_{ij} \mid x_{ij}, \beta_i \overset{iid}{\sim} \text{Bernoulli}\left(\left\{ \frac{e^{\beta_{0i}+\beta_{1i}x_{ij}}}{1 + e^{\beta_{0i}+\beta_{1i}x_{ij}}} \right\} \right) \quad \text{and}$$

$$x_{ij} \mid \nu_{\mathbf{i}} \overset{iid}{\sim} \text{Normal}\{\sum_{t=1}^{3}(\mathbf{z}'_{ij}\alpha_t + \nu_{it})w_{ijt}, \sigma_3^2\},$$

$i = 1, \ldots, \ell$, $j = 1, \ldots, N_i$. We have simulated $S = 1000$ data sets, and for each data set we fit the selection model.

Let $x_{ij}^{(s)}$, $s = 1, \ldots, S$, $i = 1, \ldots, \ell$, $j = 1, \ldots, N_i$ denote the simulated BMI values. From these values, we compute the percentiles, $Q_i^{(s)}$, by taking the appropriate ordered value. When the selection model is fit, for each s, $s = 1, \ldots, S$, the observed BMI values are $x_{ij}^{(s)} \equiv x_{ij}^{(p)}$, $i = 1, \ldots, \ell$, $j = 1, \ldots, r_i$, and the "nonresponse" BMI values $x_{ij}^{(p)}$, $i = 1, \ldots, \ell$, $j = r_i+1, \ldots, n_i$ and the "nonsampled" BMI values $x_{ij}^{(p)}$, $i = 1, \ldots, \ell$, $j = n_i + 1, \ldots, N_i$ are to be predicted. We compute the percentiles, $Q_i^{(p)}$, of the finite population, $x_{ij}^{(p)}$, $i = 1, \ldots, \ell$, $j = 1, \ldots, N_i$, by taking the appropriate ordered value. Finally, we study the relative differences

$$R_i^{(s)} = (Q_i^{(p)} - Q_i^{(s)})/Q_i^{(s)}, \quad s = 1, \ldots, S.$$

Note that although we describe the simulation for each county, prediction is done by age, race and sex for each county and appropriate adjustments to quantities like $Q_i^{(s)}$ and $Q_i^{(p)}$ are made.

Table 3: 95% credible intervals for 85th and 95th percentiles of **white male** BMI by age and model for three counties (cty)

| | | Percentile | |
| | | 85th | 95th |
cty	age	85th	95th
11	5 - 9	(17.5, 18.4)	(19.2, 20.1)
	10-14	(21.2, 22.7)	(23.2, 24.9)
	15-19	(23.1, 26.3)	(25.3, 28.8)
23	5 - 9	(18.0, 18.9)	(20.0, 21.0)
	10-14	(22.5, 24.5)	(25.0, 27.2)
	15-19	(24.8, 29.5)	(27.6, 32.7)
27	5 - 9	(18.4, 19.5)	(20.6, 21.9)
	10-14	(23.6, 25.8)	(26.5, 29.0)
	15-19	(25.9, 32.1)	(29.1, 36.0)

NOTE. The differences between the race-sex groups are small. The 95% credible intervals for age 2-4 overlap considerably on the left of those for age 5-9. The 95% credible intervals for selection model and pattern mixture model are similar.

We study four quantities in the simulation. First, the distribution of the $R_i^{(s)}$ over the S experiments for all five percentiles we studied is approximately normally distributed with mean 0.002 and standard deviation 0.027; the average of the $R_i^{(s)}$ over the S experiments for all five percentiles ranges from -0.002 to 0.003 with median of 0.002. Second, we computed the probability content of the 95% credible intervals for all age-race-sex domains by county and these numbers range from 92.2% to 96.8% with median at 94.9%. Third, we have computed the average and standard deviation of the quantities P, G and D in the deviance measure, and these quantities are 3732, 3232 and 6965 with a standard deviation 80, 88 and 159 respectively. Fourth, we have plotted DRES versus PRED for each of the S experiments, and they all look similar to the one for the selection model.

Therefore, the selection model can reproduce the BMI values and the the 95% credible intervals have coverage close to the nominal value of 95%. Thus, the selection model has good statistical properties.

4 Clustering and Survey Weights

In this section, we describe how to incorporate into the original selection model (a) a clustering effect among household members (i.e., intra-class correlation) and (b) the survey weights, thereby providing two additional models. For simplicity, we compare these models with the original selection model via inference about the parameters α_1, α_2 and α_3 to investigate the influence of age, race and sex on BMI. Design information is not available for prediction in (a).

4.1 Clustering Effect

We assess a clustering effect among household members by fitting a model which incorporates an intra-class correlation. In the selection model we add the subscript k

to r_{ij}, x_{ij}, w_{ijt} and $\mathbf{z}_{ij}$ to get r_{ijk}, x_{ijk}, w_{ijkt} and $\mathbf{z}_{ijk}$, $i = 1, \ldots, \ell$, $j = 1, \ldots, c_i$, $k = 1, \ldots, m_{ij}$. For example, x_{ijk} is the BMI value for the kth member of the jth household within the ith county. Note that $\sum_{j=1}^{c_i} m_{ij} = n_i$, the sample size from the ith county.

Then letting $d_{ijk} = \sum_{t=1}^{3} (\mathbf{z}'_{ijk}\alpha_t + \nu_{ti})w_{ijkt}$, we take

$$x_{ij} \mid \alpha, \nu, \sigma_3^2, \gamma \overset{ind}{\sim} \text{Normal}[\mathbf{d}_{ij}, \sigma_3^2\{(1 - \gamma)I_{ij} + \gamma J_{ij}\}],$$

where $0 \leq \gamma \leq 1$ is the intra-class correlation coefficient, I_{ij} is an $m_{ij} \times m_{ij}$ identity matrix and J_{ij} is an $m_{ij} \times m_{ij}$ matrix of ones. It is not possible to have an intra-class correlation coefficient for each household because the majority of households has just a single member (age 2-19). We can have one intra-cluster correlation coefficient for each county. However, we believe that this refinement is not necessary because we need to learn about the importance of the overall clustering effect.

For the response part of the model, we simply replace (r_{ij}, x_{ij}) by (r_{ijk}, x_{ijk}) keeping all parameters exactly the same as in the original selection model. All the prior specifications remain the same with the addition that $\gamma \sim$ Uniform(0, 1).

In Table 2 we can see that differences between the model with intra-class correlation and the selection model are small. There are no changes in inference about the α_t, although the intra-class correlation γ is not too small (i.e., a 95% credible interval for γ is (.38, .47)).

4.2 Inclusion of Survey Weights

Let $\tilde{\pi}_{ij}$, $j = 1, \ldots, n_i$, $i = 1, \ldots, \ell$ (calibrated such that $\sum_{j=1}^{n_i} \tilde{\pi}_{ij} = N_i$) denote the survey weights which are all known (i.e., we know the sampling weights for both respondents and nonrespondents in the sample). However, since we have only the observed BMI values, we need to impute the BMI values of the nonrespondents. (Note that, for children and adolescents, the sampling fraction for the thirty-five counties range from 0.01% to 0.10%.)

We use the original selection model to impute the BMI values of the non-respondents (i.e., the posterior means of the BMI values of the nonrespondents are substituted).

For the selection model, we have, for $i = 1, \ldots, \ell, \quad j = 1, \ldots, N_i$,

$$x_{ij} \mid \alpha, \nu, \sigma_3^2 \overset{ind}{\sim} \text{Normal}\{\sum_{t=1}^{3}(\mathbf{z}_{ij}'\alpha_t + \nu_{ti})w_{ijt}\sigma_3^2\}. \quad (7)$$

Following You and Rao (2003), we start with $x_i^* = \sum_{j=1}^{n_i} \pi_{ij}x_{ij}$, where $\pi_{ij} = \tilde{\pi}_{ij}/\sum_{j=1}^{n_i}\tilde{\pi}_{ij}$. It follows from (7) that

$$x_i^* \mid \alpha, \nu, \sigma_3^2 \overset{iid}{\sim} \text{Normal}\{\sum_{t=1}^{3}(\mathbf{z}_{it}^{*'}\alpha_t + w_{it}^*\nu_{it}), \sigma_3^2 \, \pi_i^*\} \quad (8)$$

where $w_{it}^* = \sum_{j=1}^{n_i}\pi_{ij}w_{ijt}$, $\mathbf{z}_{it}^{*'} = \sum_{j=1}^{n_i}\pi_{ij}w_{ijt}\mathbf{z}_{ij}$ and $\pi_i^* = \sum_{j=1}^{n_i}\pi_{ij}^2$. As a surrogate for x_i^*, we have used $x_i^* = \sum_{j=1}^{r_i}\pi_{ij}x_{ij} + \sum_{j=r_i+1}^{n_i}\pi_{ij}\tilde{x}_{ij}$, where $\tilde{x}_{ij}$ are the BMI values of the nonrespondents (obtained from the original selection moodel) and we assume that $r_i = \sum_{j=1}^{n_i}r_{ij}$ are fixed and known. Our prior distributions are

$$\nu_{ti} \mid \sigma_{t+3}^2 \overset{iid}{\sim} \text{Normal}(0, \sigma_{t+3}^2), \quad i = 1, \ldots, \ell,$$

$$\alpha_t \mid \sigma_{t+3}^2 \overset{ind}{\sim} \text{Normal}(\alpha_t^{(0)}, \Delta_t^{(0)})$$

$$\text{and} \quad \sigma_{t+3}^{-2} \overset{iid}{\sim} \text{Gamma}(a/2, a/2), \quad t = 1, 2, 3. \quad (9)$$

We use the same specification for a, $\alpha_t^{(0)}$ and $\Delta_t^{(0)}$ as in the selection model. Then the model, consisting of (8) and (9), is easy to fit using the Gibbs sampler. Prediction can be done in (7) after samples are taken from the posterior density obtained from (8) and (9).

In Table 2 we see that inference about α_t under the model with the survey weights is not so close to the original selection model. The loss of efficiency in estimating the regression parameters has been recognized by You and Rao (2003); so they resorted to pseudo hierarchical Bayes methods. Methods, which use estimating functions, lead to less loss in efficiency, make approximations, and the joint distribution of the sample values is still singular. In our case the large variation in the estimates is due to inefficient estimation of σ_3^2. The 95% credible interval for σ_3^2 is (46, 734) for the model with survey weights compared with (0.69, 0.77) for the original selection model (i.e., without survey weights). We have done a similar analysis using household as the unit, and got very similar results. Finally, we note that Malec, Davis and Cao (1999) use a pseudo likelihood to include sampling weights for the prevalence of overweight (binary) among adults in NHANES III with negligible nonresponse.

We have performed an exploratory data analysis on the selection probabilities (reciprocal of the sample weights) and the BMI values. A large majority of the selection probabilities is very similar: the five number summaries are .0000, .0001, .0003, .0006, .0046. A scatter plot of the BMI values versus the selection probabilities show no pattern; the correlation between the BMI values and the selection probabilities is $-.18$. Thus, it appears the selection probabilities provide little information about BMI.

References

Chib, S., and Greenberg, E. (1995), "Understanding the Metropolis-Hastings Algorithm," *The American Statistician*, **49**, 327-335.

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.

Gelfand, A., and Ghosh, S. K. (1998), "Model Choice: A Minimum Posterior Predictive Loss 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.

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: Wiley.

Malec, D., Davis, W. W., and Cao, X. (1999), "Model-Based Small Area Estimates of Overweight Prevalence Using Sample Selection Adjustment," *Statistics in Medicine*, **18**, 3189-3200.

Nandram, B. and Choi, J.W. (2005), "Hierarchical Bayesian Nonignorable Nonresponse Regression Models for Small Areas: An Application to the NHANES Data," *Survey Methodology*, **31**, 73-84.

Nandram, B., 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.

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. (1977), "Formalizing Subjective Notions About the Effect of Nonrespondents in Sample Surveys," *Journal of the American Statistical Association*, **72**, 538-543.

You, Y., and Rao, J. K. N. (2003), "Pseudo Hierarchical Bayes Small Area Estimation Combining Unit Level Models and Survey Weights," *Journal of Statistical Planning and Inference*, **111**, 197-208.