

# Parametric and Semiparametric Model-Based Estimates of the Finite Population Mean for Two-Stage Cluster Samples with Item Nonresponse

Ying Yuan<sup>1</sup> and Roderick J. A. Little<sup>2,\*</sup>

Department of Biostatistics and Applied Mathematics, M.D. Anderson Cancer Center,  
Houston, Texas 77030, U.S.A.

Department of Biostatistics, The University of Michigan, Ann Arbor, Michigan 48109, U.S.A.

\**email*: rlittle@umich.edu

**SUMMARY.** This article concerns item nonresponse adjustment for two-stage cluster samples. Specifically, we focus on two types of nonignorable nonresponse: nonresponse depending on covariates and underlying cluster characteristics, and depending on covariates and the missing outcome. In these circumstances, standard weighting and imputation adjustments are liable to be biased. To obtain consistent estimates, we extend the standard random-effects model by modeling these two types of missing data mechanism. We also propose semiparametric approaches based on fitting a spline on the propensity score, to weaken assumptions about the relationship between the outcome and covariates. These new methods are compared with existing approaches by simulation. The National Health and Nutrition Examination Survey data are used to illustrate these approaches.

**KEY WORDS:** Cluster-specific nonignorable nonresponse; Item nonresponse; Outcome-specific nonignorable nonresponse; Penalized spline of propensity prediction; Two-stage cluster sample.

## 1. Introduction

This article concerns estimation of finite population mean for cluster samples with item nonresponse. For simplicity, we restrict our attention to two-stage cluster samples. The extension of our methodology to multistage cluster samples is straightforward. We consider a population of size  $M$  consisting of  $N$  clusters with  $M_i$  elements in the  $i$ th cluster. Let  $Y_{ij}$  denote the value of a survey outcome  $Y$ , and  $X_{ij} = (X_{1ij}, \dots, X_{Pij})$  denote values of  $P$  covariates  $X = (X_1, \dots, X_P)$  for unit  $j$  in the cluster  $i$ , for  $i = 1, \dots, N; j = 1, \dots, M_i$ . Let  $T = \sum_{i=1}^N \sum_{j=1}^{M_i} Y_{ij}$  and  $\bar{Y} = T/M$  denote the finite population total and mean, respectively. At the first stage, a sample of  $n$  of the  $N$  clusters (primary sampling units, PSU's) is selected. At the second stage,  $m_i$  of the  $M_i$  units (secondary sampling units, SSU's) are selected in the  $i$ th sampled cluster, but only  $r_i$  of the  $m_i$  sampled units respond. We observe values of  $Y$  for  $r_i$  respondents, and values of  $X$  for both respondents and nonrespondents. This occurs in particular when the outcome variable  $Y$  is a sensitive question and has item nonresponse. We assume that selection probabilities for sampled units, denoted by  $\pi_{ij}$  for  $i = 1, \dots, n$  and  $j = 1, \dots, m_i$ , are known, and sampling is noninformative in the sense that  $\pi_{ij}$  is independent of  $Y_{ij}$  after conditioning on  $X$ . This assumption is often achieved by including among  $X$  the design variables that define the selection probabilities (Little, 2004). We consider estimates of the finite population mean  $\bar{Y}$  or total  $T$ . We adopt a model-based approach to multiply impute the missing values of  $Y$ , where  $Y_{ij}$ 's are clus-

tered realizations from a super-population. Having created multiply-imputed data sets, we apply standard design-based complete-data methods, specifically the Horvitz–Thompson estimate of the population mean and the design-based estimate of variance, with multiple imputation combining rules (Rubin, 1987) to incorporate imputation uncertainty. An alternative approach is to use the model for the overall inference, which (in the Bayesian paradigm) is based on the posterior predictive distribution of the population mean (e.g., Little, 2004). The choice between these two approaches involves general issues of design versus model-based inference; only the first approach is developed here in order to focus on alternative treatments of the missing data.

We apply our proposed methods to data from the Third National Health and Nutrition Examination Survey (NHANES III), conducted by the National Center for Health Statistics to assess the health of the U.S. population (National Center for Health Statistics, 2006). Given the increasing prevalence of overweight children and adolescents in the United States and its serious consequences and implications for public health (Ogden et al., 2002), a variable of great interest is body mass index (BMI) for children and youths less than 16 years old. We consider the problem of estimating the population mean of this variable, using the NHANES III youth data. This variable is subject to about 31% item nonresponse, and nonresponse is potentially nonignorable, because in health surveys the propensity to respond is often related to health (Cohen and Duffy, 2002). Methods need to account for

potential nonignorable nonresponse, and also need to account for the stratified multistage probability design of NHANES III.

One common practical approach to estimating the finite population mean  $\bar{Y}$  in the presence of item nonresponse is to use covariate information to impute the missing values of  $Y$ , and then apply standard design-based methods to the filled-in data, such as the Horvitz–Thompson estimator (Horvitz and Thompson, 1952). This method requires a model for predicting the missing values of  $Y$ . A common approach is regression (REG) imputation, where the imputed values  $\hat{y}_{ij}$  are predictions from a regression of  $Y$  on covariates  $x_{ij} = (x_{1ij}, x_{2ij}, \dots, x_{p ij})$ , with coefficients estimated from the complete data. These imputations ignore the clustering of the sample, a defect that is addressed by including random effects for clusters in the regression model, as follows:

$$[y_{ij} \mid \alpha_i, \beta_0, \beta_p, \sigma^2, \tau^2] \sim N \left( \alpha_i + \beta_0 + \sum_{p=1}^P \beta_p x_{p ij}, \sigma^2 \right),$$

$$[\alpha_i \mid \tau^2] \sim N(0, \tau^2), \tag{1}$$

where  $N(\cdot)$  denotes normal distribution. The value  $\hat{\alpha}_i$  is now included in the prediction of  $y_{ij}$ . This model (denoted RE) adds covariates to the random-effects model first proposed for cluster samples by Scott and Smith (1969).

It is well known that RE yields a consistent estimate of  $\bar{Y}$  when nonresponse is missing at random (MAR), providing the regression equation is correctly specified (Little and Rubin, 2002). Because the clusters are explicitly modeled via random effects, one might assume that RE is also valid when the missing-data mechanism depends on the clusters. However, that is not the case. When the nonresponse probability of  $y_{ij}$  depends on the cluster-specific random effect  $\alpha_i$ , the missing data mechanism is not MAR, because  $\alpha_i$ 's are not observed (Little and Rubin, 2002, Example 6.24; Yuan and Little, 2007). As in the latter paper for the case with no covariates, we use the term cluster-specific nonignorable (CNI) nonresponse to describe the mechanism that probability of response depends on underlying cluster effects  $\alpha_i$  and observed covariates, but not on observed survey outcomes within clusters.

We consider parametric models and semiparametric models that include a spline of the response propensity (Little and An, 2004) that are robust to misspecification of the mean structure relating  $Y$  to the covariates. These extend models previously proposed by Yuan and Little (2007) for the case of unit nonresponse, and yield consistent estimates of the mean under CNI nonresponse. We also consider variants of these models for the outcome-specific nonignorable (ONI) nonresponse, where missingness depends directly on the value of the outcome variable  $Y$ . Table 1 spells out all the abbreviations used in the article to label models and methods.

The next section describes our parametric and semiparametric models for estimating the finite population mean under CNI and ONI nonresponse. Section 3 describes a simulation study comparing these methods with the existing methods described above. Section 4 illustrates the methods with the NHANES III youth BMI data. Section 5 discusses our findings and conclusions.

**Table 1**  
*List of major abbreviations*

Abbreviation	
CNI	Cluster-specific nonignorable
ONI	Outcome-specific nonignorable
REG	Design-based regression
RE	Random-effects model
PCNI	Parametric cluster-specific nonignorable model
PONI	Parametric outcome-specific nonignorable model
PSPRE	Penalized spline of propensity random-effects model
PSPCNI	Penalized spline of propensity cluster-specific nonignorable model
PSPONI	Penalized spline of propensity outcome-specific nonignorable model

**2. Models**

*2.1 Parametric Models for CNI and ONI Nonresponse*

We first consider nonresponse adjustment for CNI nonresponse. In our proposed models, we model nonresponse via a normal latent variable  $Z$ , such that subject  $j$  in cluster  $i$  responds ( $r_{ij} = 1$ ) if  $z_{ij}$  is positive and fails to respond ( $r_{ij} = 0$ ) if  $z_{ij}$  is negative. As noted above, RE leads to biased estimates under CNI nonresponse. To correct this bias, we propose the following parametric cluster-specific nonignorable (PCNI) model:

$$[y_{ij} \mid \alpha_i, \chi_i, \delta, \beta_0, \beta_p, \sigma^2] \sim N \left( \alpha_i + \delta \chi_i + \beta_0 + \sum_{p=1}^P \beta_p x_{p ij}, \sigma^2 \right),$$

$$[z_{ij} \mid \chi_i, \gamma_0, \gamma_p] \sim N \left( \chi_i + \gamma_0 + \sum_{p=1}^P \gamma_p x_{p ij}, 1 \right),$$

$$[\alpha_i \mid \chi_i, \tau^2] \sim N(0, \tau^2), [\chi_i \mid \alpha_i, \omega^2] \sim N(0, \omega^2). \tag{2}$$

In this model-based approach, the survey design is taken into account by including design variables or selection probabilities as covariates. For example, if the selection probability  $\pi_{ij}$  depends on cluster sizes, and the value of  $Y$  also potentially depends on cluster sizes, we may include  $M_i$ 's as one of the covariates. In the case that design variables (e.g.,  $M_i$ 's) are unknown, we can simply include  $\pi_{ij}$  as a covariate to incorporate the selection probabilities in our model (Zheng and Little, 2004). Random effects  $\alpha_i$  and  $\chi_i$  model within-cluster correlations, and are independent of design variables (e.g.,  $M_i$  and  $m_i$ ). The PCNI model asserts that conditional on  $\chi_i$ ,  $y_{ij}$  and  $z_{ij}$ , thus  $y_{ij}$  and  $r_{ij}$ , are independent. This demonstrates an important feature of CNI nonresponse: conditional on clusters, the nonresponse probability does *not* depend on the missing values. In (2), different covariates can be used to model  $y_{ij}$  and  $z_{ij}$ , by setting some of the regression coefficients  $\{\beta_p\}$  or  $\{\gamma_p\}$  equal to zero.

Inference for the finite population mean  $\bar{Y}$  under the PCNI model can be achieved by multiple imputation (Rubin, 1987; Little and Rubin, 2002), which takes account of imputation

uncertainty. Specifically, we form  $K$  imputed data sets by filling in missing values of  $y$  with  $K$  independent draws from their posterior distribution, based on the PCNI model, with non-informative priors on the fixed parameters. These posterior draws of missing values of  $y$  can be obtained by first drawing parameters  $\{\alpha_i, \delta, \chi_i, \beta_0, \dots, \beta_p, \sigma^2\}$  from their posterior distributions, then drawing the missing values of  $y$  based on the first equation of (2). For the  $k$ th imputed data set,  $\hat{Y}$  is estimated by Horvitz–Thompson estimator as follows:

$$\hat{Y}_k = \sum_{i=1}^n \left( \sum_{j=1}^{r_i} \frac{y_{ij}}{\pi_{ij}} + \sum_{j=r_i+1}^{m_i} \frac{\hat{y}_{ij}^{(k)}}{\pi_{ij}} \right) \Big/ \sum_{i=1}^n \sum_{j=1}^{m_i} \pi_{ij}^{-1},$$

where  $\hat{y}_{ij}^{(k)}$  is the  $k$ th imputed value of  $y_{ij}$ . Then, a consistent estimate of  $\bar{Y}$  is given by  $\hat{\bar{Y}} = \frac{1}{K} \sum_{k=1}^K \hat{Y}_k$ , and its variance is  $\text{Var}(\hat{\bar{Y}}) = \frac{1}{K} \sum_{k=1}^K V_k + \frac{1}{K-1} \sum_{k=1}^K (\hat{Y}_k - \hat{\bar{Y}})^2$ , where  $V_k$  is the variance of  $\hat{Y}_k$  estimated from the  $k$ th imputed data set. Note that the super-population model (2) is used to predict the missing values of  $y$ . Once the missing data are filled in by model predictions, the design-based weighting method is used to estimate the finite population mean. This approach limits the effects of model misspecification to the treatment of the missing data. An alternative approach, potentially more efficient because it makes fuller use of the super-population model but also more vulnerable to model misspecification, is to base the inferences on the posterior distribution of the population mean. We do not consider that approach here in order to focus attention on the missing data adjustments.

The above model assumes a CNI mechanism, where  $r_{ij}$  and  $y_{ij}$  are independent within clusters (i.e., conditional on  $\chi_i$ ) after conditioning on covariates  $x_{ij}$ . An ONI mechanism assumes that missingness of  $y_{ij}$  depends directly on the value of  $y_{ij}$  and observed covariates. Such a mechanism is modeled by the following PONI model, where  $y_{ij}$  has a linear regression on  $z_{ij}$  rather than the random effect  $\chi_i$ :

$$\begin{aligned} [y_{ij} \mid z_{ij}, \alpha_i, \delta, \beta_0, \beta_p, \sigma^2] &\sim N \left( \alpha_i + \delta z_{ij} + \beta_0 + \sum_{p=1}^P \beta_p x_{p ij}, \sigma^2 \right), \\ [z_{ij} \mid \chi_i, \gamma_0, \gamma_p] &\sim N \left( \chi_i + \gamma_0 + \sum_{p=1}^P \gamma_p x_{p ij}, 1 \right), \\ [\alpha_i \mid \chi_i, \tau^2] &\sim N(0, \tau^2), [\chi_i \mid \alpha_i, \omega^2] = N(0, \omega^2). \end{aligned}$$

2.2 Propensity Spline Models

The foregoing models assume the correct parametric specification of  $E(y_{ij} \mid x_{ij})$ . A more robust approach is to model  $E(y_{ij} \mid x_{ij})$  nonparametrically, replacing the linear predictor  $\sum_{p=1}^P \beta_p x_{p ij}$  by an arbitrary nonparametric function  $h(x_{1 ij}, \dots, x_{P ij})$ . With a single covariate,  $h(\cdot)$  can be easily estimated by a kernel or spline regression. However, as the number of covariates  $P$  becomes large, the estimation of  $h(\cdot)$  becomes difficult, the so-called ‘‘curse of dimensionality.’’ To address this problem, Little and An (2004) proposed penalized spline of propensity prediction (PSPP), which yields a doubly robust estimate by focusing the spline on a particular function of the covariates most sensitive to model misspecification,

namely the estimated propensity score. They studied the situation where observations are independent and the missing data mechanism is ignorable. We extend the PSPP approach to correlated two-stage cluster samples with ignorable, CNI, or ONI nonresponse.

We first consider ignorable or MAR nonresponse where missingness of  $y_{ij}$  only depends on the covariates  $x_{ij}$ . Let  $v_{ij}$  denote the probit of the propensity score for the subject  $j$  in the cluster  $i$ , and we assume for convenience the standard probit model for nonresponse as follows:

$$v_{ij} = \Phi^{-1}(\text{Pr}(r_{ij} = 1 \mid x_{ij})) = \gamma_0 + \gamma_1 x_{1 ij} + \dots + \gamma_P x_{P ij}. \tag{3}$$

Conditional on the propensity score, the missingness of  $y_{ij}$  does not depend on  $x_{ij}$  (Rosenbaum and Rubin, 1983), namely,  $\text{Pr}(r_{ij} = 1 \mid x_{ij}, v_{ij}) = \text{Pr}(r_{ij} = 1 \mid v_{ij})$ . This fact motivates the following penalized spline of propensity random-effects (PSPRE) model:

$$\begin{aligned} [y_{ij} \mid \alpha_i, \beta_p^*, \theta, \sigma^2] &\sim N \left( \alpha_i + S(v_{ij}; \theta) + \sum_{p=1}^{P-1} \beta_p^* x_{p ij}^*, \sigma^2 \right), \\ [\alpha_i \mid \tau^2] &\sim N(0, \tau^2), \end{aligned} \tag{4}$$

where  $S(v_{ij}; \theta)$  denotes a spline regression on  $v_{ij}$  with parameters  $\theta$ , and  $x_{p ij}^*$  is defined as the residual of  $x_{p ij}$  from the following spline regression indexed by parameters  $\theta_p$ :

$$[x_{p ij} \mid v_{ij}, \theta_p, \sigma_p^2] \sim N(S_p(v_{ij}; \theta_p), \sigma_p^2).$$

In this case,  $x_{p ij}^*$  is given by

$$\begin{aligned} x_{p ij}^* &= x_{p ij} - E(x_{p ij} \mid v_{ij}) \\ &= x_{p ij} - S_p(v_{ij}; \theta_p), \quad p = 1, \dots, P - 1. \end{aligned} \tag{5}$$

In the PSPRE model, the regression on  $\{x_{p ij}^*; p = 1, \dots, P - 1\}$  is not necessary for the consistency of the estimate, but it improves efficiency if  $x_{ij}$  is a good predictor of  $y_{ij}$ . To avoid multicollinearity,  $x_{p ij}^*$  is dropped from (4). In general, to reduce colinearity, it is preferable to drop the residual of a covariate that is highly correlated with the propensity model to respond (3). To obtain an estimate of  $\bar{Y}$ , we can fit the PSPRE model to multiply-impute missing values of  $Y$ , and then apply MI procedures discussed in the Section 2.1.

For CNI nonresponse, we modify the parametric PCNI model by including a spline on the propensity score to yield the following penalized spline of propensity cluster-specific nonignorable (PSPCNI) model:

$$\begin{aligned} [y_{ij} \mid \alpha_i, \chi_i, \delta, \theta, \beta_p^*, \sigma^2] &\sim N \left( \alpha_i + \delta \chi_i + S(v_{ij}; \theta) + \sum_{p=1}^{P-1} \beta_p^* x_{p ij}^*, \sigma^2 \right), \\ [z_{ij} \mid \chi_i, \gamma_0, \gamma_p] &\sim N \left( \chi_i + \gamma_0 + \sum_{p=1}^P \gamma_p x_{p ij}, 1 \right), \\ [\alpha_i \mid \chi_i, \tau^2] &\sim N(0, \tau^2), [\chi_i \mid \alpha_i, \omega^2] \sim N(0, \omega^2), \end{aligned} \tag{6}$$

where  $v_{ij}$  is the probit of the propensity score defined as (3). Strictly, integrating out the latent variable  $z_{ij}$  in the PSPCNI model (6) yields the probit of the propensity score of the form  $v_{ij}^* = \chi_i + \gamma_0 + \gamma_1 x_{1 ij} + \dots + \gamma_P x_{P ij}$ , which includes the

unobserved random effect  $\chi_i$ . However, it is very difficult to fit splines on  $v_{ij}^*$  directly, and to simplify the estimation we model with a spline the propensity score  $v_{ij}$  that omits the term  $\chi_i$ . We do not expect this modification to degrade the properties of the method much, because the rationale behind the PSPP method is to be parametric in the direction of covariate space where respondents and nonrespondents are balanced in terms of the covariate distributions. This conjecture is supported by the simulation study in Section 3. In the above PSCNI model,  $x_{pij}^*$  is the residual defined by (5) as before.

For ONI nonresponse, we add a spline to the PONI model to obtain the following penalized spline of propensity outcome-specific nonignorable model (PSPONI):

$$\begin{aligned} & [y_{ij} \mid z_{ij}, \alpha_i, \delta, \theta, \beta_p^*, \sigma^2] \\ & \sim N \left( \alpha_i + \delta z_{ij} + S(v_{ij}; \theta) + \sum_{p=1}^{P-1} \beta_p^* x_{pij}^*, \sigma^2 \right) \\ [z_{ij} \mid \chi_i, \gamma_0, \gamma_p] & \sim N \left( \chi_i + \gamma_0 + \sum_{p=1}^P \gamma_p x_{pij}, 1 \right) \\ [\alpha_i \mid \chi_i, \tau^2] & \sim N(0, \tau^2), [\chi_i \mid \alpha_i, \omega^2] \sim N(0, \omega^2). \end{aligned}$$

These PSPP models can be easily extended to situations where some covariates are categorical. Unlike continuous covariates, categorical covariates are directly included in the mean regression model of  $y$  (without taking residuals), i.e., regressing  $y$  on  $x_{pij}$  instead of  $x_{pij}^*$ .

### 2.3 Fitting the Semiparametric Models

A convenient approach to fitting the models described in Sections 2.1 and 2.2 is to add noninformative priors for the fixed parameters and simulate draws from the posterior distribution of the parameters. Given posterior draws of these parameters, it is straightforward to draw the value of  $y$  for nonrespondents from its posterior distribution to form multiply-imputed data sets, and obtain the MI estimate of the population mean. For reviews of the Bayesian approach to sample surveys, see for example Little (2003, 2004). Estimates for the RE model are easily obtained by the Gibbs' sampler discussed in Gelfand et al. (1990). Specifically, for the RE model, we use noninformative priors for  $\beta_0, \dots, \beta_P$  and diffuse inverse gamma priors for  $\sigma^2$  and  $\tau^2$ , namely:

$$[\beta_0, \dots, \beta_P, \sigma^2, \tau^2] \propto (\sigma^2)^{-(a_1+1)} e^{-b_1/\sigma^2} (\tau^2)^{-(a_2+1)} e^{-b_2/\tau^2},$$

with  $a_1 = b_1 = a_2 = b_2 = 0.1$ , a value small enough that the information in the data strongly dominates the information in the prior distribution. For the PCNI and PONI models, we assume priors for the fixed parameters of the form

$$\begin{aligned} & [\delta, \beta_0, \dots, \beta_P, \gamma_0, \dots, \gamma_P, \sigma^2, \tau^2, \omega^2] \\ & \propto (\sigma^2)^{-(a_1+1)} e^{-b_1/\sigma^2} (\tau^2)^{-(a_2+1)} e^{-b_2/\tau^2} (\omega^2)^{-(a_3+1)} e^{-b_3/\omega^2}, \end{aligned}$$

with  $a_1 = b_1 = a_2 = b_2 = a_3 = b_3 = 0.1$ , again so that the information in the data strongly dominates the information in the prior distribution. Computation via the Gibbs' sampler is straightforward, providing the latent values  $\{z_{ij}\}$  are included as missing data.

A convenient way to implement the spline regressions for the PSPRE, PSCNI, and PSPONI models is the penalized spline method of Ruppert, Wand, and Carroll (2003). Let  $\kappa_1, \dots, \kappa_K$  denote prespecified  $K$  knots, and use linear spline base functions, the spline  $S(v_{ij})$  in PSCNI or PSPONI can be written as a random-effects model:

$$S(v_{ij}) = \beta_0 + \beta_1 v_{ij} + \sum_{k=1}^K b_k (v_{ij} - \kappa_k)_+, \quad (7)$$

$$[b_k \mid \varphi^2] = N(0, \varphi^2),$$

where  $(v_{ij} - \kappa_k)_+$  equals  $v_{ij} - \kappa_k$  if  $v_{ij} - \kappa_k > 0$ , and 0 otherwise. The tuning parameter  $\varphi^2$  is treated as a parameter and estimated from the data. By plugging in the above expressions, PSCNI and PSPONI models have similar forms to PCNI and PONI, and can be fitted by the Gibbs' sampler, as outlined for the PSPONI model in Web Appendix A. The linear spline is used here for simplicity, but other bases, such as the cubic spline basis, are also possible. Because our goal is prediction, the exact form of basis does not make much difference to the results, provided that the knots cover the range of the data reasonably well (Ruppert et al., 2003). We monitor convergence of the Gibbs' chains by graphical inspection, and by the method of Gelman and Rubin (1992) based on multiple Gibbs' chains with overdispersed starting points. Gelman and Rubin's method calculates the between-chain variance of simulated draws from multiple Gibbs' chains, and compares this with the within-chain variance. If the ratio of these two variances is near 1 (e.g., less than 1.2), we conclude that the Gibbs' sequences have converged.

### 3. Simulation Study

We now describe a simulation study to assess the performance of the above methods. We construct 16 populations of  $M = \sum_{i=1}^N M_i = 40,492$  units each, arranged in 200 clusters. Cluster sizes  $\{M_i\}$  are randomly generated from a uniform distribution between 20 and 400. Two covariates  $x_1$  and  $x_2$  are simulated as independent normal with mean 2 and standard deviation 1. The outcomes  $\{y_{ij}\}$  and response indicators  $\{r_{ij}\}$  are generated according to the model:

$$\begin{aligned} & [y_{ij} \mid z_{ij}, \alpha_i, \beta, \chi_i, \lambda, \sigma^2] \\ & \sim N(\alpha_i + \lambda \beta (z_{ij} - \chi_i) + \beta \chi_i + g_1(x_{1ij}, x_{2ij}), \sigma^2), \end{aligned}$$

$$[z_{ij} \mid \chi_i] \sim N(\chi_i + g_2(x_{1ij}, x_{2ij}), 1),$$

$$[\alpha_i \mid \chi_i, \tau^2] \sim N(0, \tau^2), [\chi_i \mid \alpha_i, \omega^2] \sim N(0, \omega^2).$$

with (a)  $r_{ij} = 1$  if  $z_{ij} > 0$ , as above; (b)  $\sigma^2 = 1$ ,  $\tau^2 = 1$  and  $\omega^2 = 4$ ; (c) two choices of mean models, linear:  $g_1(x_1, x_2) = a + x_1 + x_2$  and cubic:  $g_1(x_1, x_2) = b + 0.5x_1^3 + 0.5x_2^3$ , with  $a$  and  $b$  chosen so that the super-population mean is 20; (d) two choices of propensity models, both with an overall response rate of 0.6, namely linear:  $g_2(x_1, x_2) = x_1 + x_2 - 3.2$ , and cubic:  $g_2(x_1, x_2) = 2.06 - 0.05x_1^3 - 0.05x_2^3$ ; (e) two choices of  $\beta$ ,  $\beta = 0$  for MAR nonresponse and  $\beta = 10$  for non-MAR (NMAR) nonresponse; and (f) three choices of  $\lambda$ ,  $\lambda = 1$ , yielding an ONI mechanism where missingness of  $y_{ij}$  depends on  $y_{ij}$ ,  $x_1$  and  $x_2$ ;  $\lambda = 0$ , yielding a CNI mechanism where missingness of  $y_{ij}$  depends on  $\chi_i$ ,  $x_1$  and  $x_2$ ; and  $\lambda = 0.5$ , yielding a mixture of CNI and ONI nonresponse.

Five hundred two-stage samples were selected from each population, with  $n = 20$  PSUs (or clusters), chosen randomly with probability proportional to size, and  $m = 10$  units randomly selected from each sampled PSU, yielding a total sample size of 200 selected with equal selection probability. On each sample we compute the mean of  $Y$  before deletion (BD) and seven estimators of  $\bar{Y}$  (REG, RE, PCNI, PONI, PSPRE, PSPCNI, PSPONI) from the incomplete data. We choose 19 equally spaced knots over the estimated response propensity for the spline regressions, and conduct  $K = 20$  multiple imputations.

### 3.1 Results

Table 2(a) to (d) shows various summary statistics of seven approaches (REG, RE, PCNI, PONI, PSPRE, PSPCNI, PSPONI) over the 500 samples when missing data are MAR, CNI, ONI, and mixture of CNI and ONI under the two mean models and two propensity models. The summary statistics include empirical bias, root mean square error (RMSE) and 95% confidence interval coverage rate (nominally, we expect 0.95). We denote the estimates of  $\bar{Y}$  based on the seven models by  $\hat{Y}_{\text{REG}}$ ,  $\hat{Y}_{\text{RE}}$ ,  $\hat{Y}_{\text{PCNI}}$ , and so on.

**3.1.1 Missing at random nonresponse ( $\beta = \lambda = 0$ ).** In this case, the missingness of  $y_{ij}$  only depends on covariates  $x_{ij}$ , and the missing data mechanism is MAR. As shown in Table 2a, the performances of parametric approaches RE, PCNI, PONI depend on the linearity in the relationship between  $y$  and  $x$ . If the mean model is linear, these methods yield consistent estimates; otherwise, they lead to biased estimates. In contrast, the semiparametric approaches PSPRE, PSPCNI, and PSPONI are robust to the nonlinearity in the relationship between  $y$  and  $x$ , given the correctly specified propensity model. For example, when the true mean model is cubic and true propensity model is linear, empirical biases of  $\hat{Y}_{\text{RE}}$  and  $\hat{Y}_{\text{PCNI}}$  are  $-1.07$  and  $-1.26$ , while the empirical biases of  $\hat{Y}_{\text{PSPRE}}$  and  $\hat{Y}_{\text{PSPCNI}}$  are  $-0.17$  and  $-0.2$ . If the mean model is correctly specified, both parametric and semiparametric approaches show the robustness to the misspecification of the propensity model. Specifically, the robustness of RE is due to the fact that under MAR, RE yields consistent estimates if the mean model is correctly specified. For PSPCNI and PSPONI, because the propensity model is specified as a linear function of  $x$ , the regression on cluster-level random effects  $\chi_i$  or the latent variable  $z$  in the mean model translates into a linear regression on  $x$ , reflecting the true relationship between  $y$  and  $x$ . For semiparametric approaches, the robustness may be because covariates enter the mean model linearly via the propensity score, and as a result approximate models the linear relationship between  $y$  and  $x$ . As one might expect, if both the propensity model and the mean model are misspecified, all the methods yield bias estimates.

When nonresponse is MAR, the efficiency loss of the models assuming ONI and CNI is rather minor. For the models correctly assuming MAR, the semiparametric model PSPRE has very similar efficiency to the parametric model RE when both mean and propensity are correctly specified (i.e., linear). The high efficiency of the semiparametric model is due to the regression on residuals of covariates  $x_{pij}^*$ 's, which efficiently picks up the information ignored by the spline on propensity.

Actually, when the true mean model is linear, the linear regression on  $x_{pij}^*$ 's in the semiparametric models is equivalent to modeling the covariates parameterically, leading to highly efficient estimates. We observed the similar high efficiency for CNI and ONI semiparametric models as well.

**3.1.2 Cluster-specific nonignorable nonresponse ( $\beta = 10$ ,  $\lambda = 0$ ).** In this case, nonresponse is associated with unobserved cluster characteristics. In Table 2b, it is seen that  $\hat{Y}_{\text{RE}}$  and  $\hat{Y}_{\text{PSPRE}}$  have noticeable empirical biases of 1.35 and 1.48 even though the mean model is correctly specified, because they do not adjust for CNI nonresponse. In contrast, PCNI and PSPCNI correct this problem, yielding small empirical biases and good confidence coverage when the model is correctly specified. Again, we find that semiparametric model PSPCNI has very similar efficiency to the parametric model PCNI. These methods also do well when the propensity model is misspecified but the mean model is correctly specified. However, PCNI is sensitive to the misspecification of the mean model. PSPCNI addresses this problem, and is robust to the nonlinearity of the mean structure. For example, when the true mean model is cubic and the true propensity model is linear, the empirical bias of  $\hat{Y}_{\text{PCNI}}$  ( $-1.09$ ) is about four times as large as that of  $\hat{Y}_{\text{PSPCNI}}$  ( $-0.26$ ). Even in the case where both  $\hat{Y}_{\text{PCNI}}$  and  $\hat{Y}_{\text{PSPCNI}}$  are biased, such as when both the mean model and the propensity model are misspecified, the estimates based on PSPCNI still have a smaller bias and a better coverage rate than PCNI. The two ONI methods, PONI and PSPONI, generally lead to biased estimates for CNI nonresponse, although PSPONI does better in terms of bias and coverage.

**3.1.3 Outcome-specific nonignorable nonresponse ( $\beta = 10$ ,  $\lambda = 1$ ).** As shown in Table 2c, REG, RE, PSPRE, PCNI, PSPCNI in general lead to biased estimates and have poor coverage rates because they all misspecify the missing data mechanism. PONI and PSPONI yield consistent estimates when both the mean model and the propensity model are correctly specified. However, if the relationship between  $y$  and  $x$  is misspecified, PONI leads to biased estimates. PSPONI corrects the bias, provided that the propensity model is correctly specified. Both PONI and PSPONI lead to slightly biased estimates when the mean model is correctly specified but the propensity model is misspecified. When both the mean model and the propensity model are misspecified, the biases of PONI and PSPONI are surprisingly small, perhaps because effects of the misspecification of mean model and propensity model coincidentally canceled out.

**3.1.4 Mixed outcome and cluster-specific nonignorable nonresponse ( $\beta = 10$ ,  $\lambda = 0.5$ ).** This nonignorable missing data mechanism is a mixture of the CNI and the ONI nonresponse. None of the methods is satisfactory in terms of bias, RMSE, and coverage, although PONI and PCNI do well in certain situations, such as when both mean and propensity are linear. There is no single method that dominates the others consistently over the simulation conditions.

## 4. Application

We apply our methods to the NHANES III youth data for children and youths 2 months to 16 years of age. The survey design of NHANES is a stratified multistage probability

**Table 2**

Empirical bias ( $\times 100$ ), RMSE ( $\times 100$ ), and coverage rate (Cov) of 95% confidence intervals (%) for seven methods when nonresponse is (a) MAR, (b) CNI, (c) ONI, and (d) the mixture of CNI and ONI, under different mean and propensity models

Mean	Propensity		BD	REG	RE	PSPRE	PCNI	PSPCNI	PONI	PSPONI
(a) MAR ( $\beta = \lambda = 0$ )										
Linear	Linear	Bias	-1	-4	-3	-3	-2	-2	0	-4
		RMSE	25	29	28	28	30	30	37	43
		Cov	95.4	95.0	95.6	95.6	95.4	96.2	96.4	95.8
Linear	Cubic	Bias	-1	-6	-7	-7	-6	-6	-8	-14
		RMSE	25	29	28	27	31	31	47	64
		Cov	95.4	95.4	96.2	95.6	97.0	96.6	96.0	96.6
Cubic	Linear	Bias	-1	-106	-107	-17	-126	-20	-393	-29
		RMSE	90	151	153	99	179	110	505	144
		Cov	96.4	84.4	89.0	97.6	86.0	97.6	46.8	96.6
Cubic	Cubic	Bias	-1	-166	-172	-106	-213	-123	-475	-164
		RMSE	90	189	194	138	231	152	522	202
		Cov	96.4	54.8	39.8	71.0	34.2	71.0	18.0	70.8
(b) CNI ( $\beta = 10, \lambda = 0$ )										
Linear	Linear	Bias	13	1080	135	148	6	15	114	84
		RMSE	429	1153	421	440	459	456	420	422
		Cov	95.4	25.0	93.4	93.2	95.4	95.4	94.6	96.2
Linear	Cubic	Bias	13	1027	196	228	-15	-37	166	117
		RMSE	429	1105	451	466	474	471	438	418
		Cov	95.4	32.0	92.8	91.4	95.4	95.4	93.8	95.2
Cubic	Linear	Bias	13	977	148	211	-109	-26	-378	-172
		RMSE	434	1069	450	466	491	468	654	525
		Cov	95.2	44.0	93.4	93.8	93.6	95.4	86.8	92.6
Cubic	Cubic	Bias	13	867	151	213	-263	-173	-292	-2
		RMSE	434	974	479	493	593	522	564	448
		Cov	95.2	56.0	92.8	92.4	92.0	94.0	89.4	94.6
(c) ONI ( $\beta = 10, \lambda = 1$ )										
Linear	Linear	Bias	10	1385	779	784	499	503	-15	12
		RMSE	450	1430	838	843	622	622	476	480
		Cov	94.6	0.8	43.2	44.2	77.2	77.2	95.6	95.4
Linear	Cubic	Bias	12	1493	1088	1063	701	666	79	90
		RMSE	450	1530	1128	1103	776	745	473	472
		Cov	94.6	0.0	3.2	4.8	53.4	59.8	95.8	95.6
Cubic	Linear	Bias	10	1283	718	787	386	483	-261	-81
		RMSE	468	1344	800	855	574	629	621	549
		Cov	94.6	9.4	60.6	52.2	87.6	79.0	93.2	94.6
Cubic	Cubic	Bias	12	1333	899	909	505	518	-77	-40
		RMSE	439	1380	953	961	614	625	494	486
		Cov	95.2	2.8	18.6	17.4	76.2	75.2	95.6	95.8
(d) Mixture of ONI and ONI ( $\beta = 10, \lambda = 0.5$ )										
Linear	Linear	Bias	12	1232	463	466	249	249	48	52
		RMSE	436	1288	582	587	472	474	435	438
		Cov	95.0	8.4	81.8	81.4	92.0	92.4	94.6	95.2
Linear	Cubic	Bias	13	1260	683	668	328	314	137	141
		RMSE	436	1312	770	756	510	501	441	441
		Cov	94.6	5.8	55.4	58.6	88.0	89.2	95.2	95.6
Cubic	Linear	Bias	13	1625	1006	839	949	774	831	489
		RMSE	471	1674	1072	920	1023	866	977	656
		Cov	94.0	1.2	30.4	52.0	38.0	60.2	52.8	83.4
Cubic	Cubic	Bias	13	1100	508	540	133	180	-80	18
		RMSE	433	1169	635	658	438	451	464	440
		Cov	94.8	21.2	75.6	74.8	94.4	93.8	94.2	95.2

**Table 3**

*Estimates of the finite population mean of BMI, weight in kilograms divided by the square of height in meters, for U.S. children and youths 2 months to 16 years of age.  $\hat{Y}$  is the estimate of the BMI, and  $SE(\hat{Y})$  is the estimate of the associated standard error.*

Estimates	Methods						
	REG	RE	PSPRE	PCNI	PSPCNI	PONI	PSPONI
$\hat{Y}$	17.85	17.84	17.76	17.84	17.72	17.05	16.82
$SE(\hat{Y})$	0.34	0.33	0.34	0.33	0.33	0.31	0.32

sample of the civilian noninstitutionalized U.S. population with counties as PSUs (National Center for Health Statistics, 1994). The sample was selected from households in 81 counties across the United States during the period from October 1988 to October 1994. For confidentiality reasons, county identifiers for counties with population less than 500,000 are suppressed. As a result, the public use data we analyzed here came from only the 35 largest counties with population at least 500,000. In our analysis, we considered clustering at the county level, although there may be a need to include clustering at the household level. Nandram, Han, and Choi (2002) and Nandram and Choi (2005) reported that the clustering at the household level is negligible and can be ignored as a first approximation. Following their approach, we ignore the clustering within households, i.e., approximating multilevel clustering by two-level clustering.

The NHANES III consists of two components. The first component is the interview of the sampled individuals by a NHANES interviewer to collect personal information, such as demographic variables. The second component is the health examination of interviewed persons at a mobile examination center or their home to collect health data. BMI data are collected at the health examination.

Nonresponse occurs in both the interview and examination components of the survey. The interview nonresponse arises when the sampled individuals did not respond for the interview. The examination nonresponse arises when the interviewed subjects failed to arrive at the mobile examination center for health examinations or missed examinations at home, thereby missing all or part of the examinations. In NHANES III youth data for children and youths 2 months to 16 years of age, BMI have 31% missing values. In contrast, sex, race, age, years of education, household income, mother’s BMI, and father’s BMI have much lower nonresponse rates, and are used as covariates to predict the missing values of income. Here we do not consider the small number of individuals

whose BMI values and the covariates are missing (i.e., unit nonresponse), but only include individuals with complete covariates (i.e., complete cases and item nonresponse) in our data analysis. All the covariates except sex and race are treated as continuous variables in our models. Sex and race (white or nonwhite) are binary variables. To improve the normality, a logarithm transformation is applied to BMI variables and a cube root transformation is applied to household income. 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.” The nonresponse of BMI in NHANES III may not be random and are likely to be nonignorable, motivating us to apply our methods.

Table 3 shows estimates of the finite population mean of BMI for children and youths 2 months to 16 years of age and associated standard errors under different models. In general, all methods except the ONI models PONI and PSPONI yield similar results, that is, the average BMI is about 17.80, in the unit of kilogram/meter<sup>2</sup> (kg/m<sup>2</sup>). The estimates of two ONI models are slightly lower than the other models. The estimate of average BMI based on the POSNI is 17.05 kg/m<sup>2</sup>, and the estimate of average BMI based on the PSPONI is 16.82 kg/m<sup>2</sup>. The similarity between the results from these two ONI models may be partly due to the fact that the relationships between the BMI and the covariates do not severely deviate from linearity. Nandram and Choi (2005) obtained similar results when comparing a linear model and a spline regression model for NHANES III BMI data, although they emphasized small area estimation.

Table 4 displays estimates of selected parameters for different models. Estimates of residual variance  $\sigma^2$  based on PONI and PSPONI are significantly smaller than the other models, suggesting that modeling ONI missing data mechanism substantially decreases the residual variance. However, estimates of random intercept variance  $\tau^2$  are similar for all models. For nonignorable models PCNI, PSPCNI, PONI, and PSPONI,

**Table 4**

*Estimates of selected parameters under different models based on NHANES III youth BMI data*

Parameters	Models						
	REG	RE	PSPRE	PCNI	PSPCNI	PONI	PSPONI
$\sigma^2$	0.0222	0.0220	0.0207	0.0221	0.0220	0.0026	0.0023
$\tau^2$	N/A	0.0039	0.0036	0.0041	0.0040	0.0036	0.0031
$\delta$	N/A	N/A	N/A	0.0302	0.0295	0.1655	0.1652
$\omega^2$	N/A	N/A	N/A	0.0301	0.0299	0.0257	0.0281

giving the similar values of  $\hat{\omega}^2$ , the parameter  $\delta$  is of particular interest because it determines the nonignorability. The value  $\hat{\delta}$  are rather small for PCNI and PSPCNI, consistent with the finding that  $\hat{Y}$  based on PCNI and PSPCNI are very similar to ignorable models. In contrast,  $\hat{\delta}$  is relatively much larger for PONI and PSPONI, suggesting substantial nonignorability of the missing data, and also explaining the difference between  $\hat{Y}_{\text{PONI}}$ ,  $\hat{Y}_{\text{PSPONI}}$ , and other models.

From the statistical point of view, there is no information to identify the ONI nonresponse based on observed data. However, the comparison among ignorable, CNI, and ONI models provides a form of sensitivity analysis for the missing data mechanisms of nonresponse. For this particular data, we may present the result as that the average BMI for youths less than 16 years old is about 16.8 to 17.8 kg/m<sup>2</sup>. We note that one reason for the nonresponse is that parents of very young children (e.g., less than 2 years old) are often protective and would not allow their children to leave home for a physical examination. Because the very young children are less likely to be overweight, the nonresponse here may be ONI in the sense that individuals with lower BMI are less likely to respond. However, by conditioning on variables that are highly related to BMI, the bias from ONI nonresponse may be substantially reduced.

## 5. Conclusion

This article studies item nonresponse adjustment for two-stage cluster samples. When nonresponse is MAR, the parametric approaches RE, PCNI, and PONI yield consistent estimates when the mean model is correctly specified. However, if the mean model is misspecified, they lead to biased estimates. Without imposing any parametric form on the mean structure, the semiparametric approaches PSPRE, PSPONI, and PSPCNI are robust to the nonlinearity between the outcome and covariates. When nonresponse is CNI, RE and PSPRE lead to biased estimates because the MAR assumption is violated. PCNI takes account of CNI nonresponse and yields consistent estimates if the mean structure is correctly specified, but otherwise is potentially biased. PSPCNI addresses this problem, and yields consistent estimates under nonlinear mean structures. Under CNI nonresponse, PONI and PSPONI generally lead to biased estimates although PSPONI does better in terms of bias and coverage than PONI, suggesting that estimates based on PONI and PSPONI are sensitive to the misspecification of the nonignorable missing data mechanism. When nonresponse is ONI, all methods except PONI and PSPONI are biased and have poor coverage rates. PSPONI is more robust than PONI in the sense that PSPONI does not require the linearity between the outcome and covariates. When the nonresponse is a mixture of CNI and ONI, none of the methods is satisfactory.

A natural and important question given these findings is if we can determine whether nonresponse is CNI or ONI, and which model should be used. Unfortunately, we cannot distinguish between these two types of nonignorable nonresponses solely based on the observed data. Nevertheless, if auxiliary variables for nonrespondents and respondents are available, for example from census data, then we could compare the residual distribution of auxiliary variables, obtained by regressing on appropriate covariates, of nonrespondents with

that of respondents within a cluster. If there is no systematic difference, we might assume that nonresponse is more likely to depend on underlying cluster-specific characteristics and apply PCNI or PSPCNI; otherwise, we may consider PONI or PSPONI. Because we do not have enough information to distinguish between alternative nonignorable missing mechanisms, it may be more appropriate to apply more than one method and compare results. In addition, the cognitive and social psychological theory of survey participation also provides an important way to understand and identify the nonignorable nonresponse mechanism (Groves and Couper, 1998).

We focus on two-stage cluster samples, but our models can be readily extended to accommodate multistage cluster samples. To reflect the design feature of multistage cluster sampling, we can use multilevel hierarchical models where different levels within-cluster correlations are modeled by different level random effects. The same latent variable approach can be used to model the ONI missing data mechanism. CNI nonresponse now may associate with random effects of different levels. In principle, we could extend PCNI and PSPCNI model to allow the mean of the outcome variable to depend on random effects of all levels. However, because substantial within-cluster correlation may only occur in certain levels, it may be adequate to just model the relationship between the outcome variables and the random effects of these levels.

In the propensity spline models discussed in Section 2.2, the covariates  $x_{pij}$  were replaced by centered versions  $x_{pij}^*$  as in Eq. (5). Recent work in Zhang and Little (2005) suggests that this centering is not necessary to yield robust estimates. This simplifies the model.

This article concerns estimation of the finite population means. In this case, it is more critical to correctly specify the mean model rather than the propensity model. For this reason and also for the sake of convenience, we assume the parametric linear probit model for nonresponse; that is, the mean of the latent variable  $z_{ij}$  depends on covariates  $x_{ij}$  linearly. Of course, to obtain more robust approaches, we could model the relationship between  $z_{ij}$  and  $x_{ij}$  nonparametrically as well, for example, using generalized additive models (Hastie and Tibshirani, 1990).

## ACKNOWLEDGEMENTS

The authors thank the associate editor for the insightful and detailed comments that greatly improved this article.

## REFERENCES

- Cohen, G. and Duffy, J. C. (2002). Are nonrespondents to health surveys less healthy than respondents. *Journal of Official Statistics* **18**, 13–23.
- Gelfand, A. E., Hills, S. E., Racine-Poon, A., and Smith, A. F. M. (1990). Illustration of Bayesian inference in normal data models using Gibbs sampling. *Journal of the American Statistical Association* **85**, 972–985.
- Gelman, A. and Rubin, D. (1992). Inference from iterative simulation using multiple sequences. *Statistical Science* **7**, 457–511.
- Groves, R. M. and Couper, M. P. (1998). *Nonresponse in Household Interview in Surveys*. New York: Wiley.
- Hastie, T. J. and Tibshirani, R. J. (1990). *Generalized Additive Models*. New York: Chapman & Hall.



- Horvitz, D. G. and Thompson, D. J. (1952). A generalization of sampling without replacement from a finite population. *Journal of the American Statistical Association* **47**, 663–685.
- Little, R. J. A. (2003). The Bayesian approach to sample survey inference. In *Analysis of Survey Data*, R. L. Chambers and C. J. Skinner (eds), 49–57. New York: John Wiley & Sons.
- Little, R. J. A. (2004). To model or not to model? Competing models of inference for finite population sampling. *Journal of the American Statistical Association* **99**, 546–556.
- Little, R. J. A. and An, H. (2004). Robust likelihood-based analysis of multivariate data with missing values. *Statistica Sinica* **14**, 949–968.
- Little, R. J. A. and Rubin, D. B. (2002). *Statistical Analysis with Missing Data*, 2nd edition. New York: John Wiley & Sons.
- Nandram, B. and Choi, J. W. (2005). Hierarchical Bayesian nonignorable 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.
- 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.
- National Center for Health Statistics. (2006). Available at <http://www.cdc.gov/nchs/nhanes.htm>.
- Ogden, C. L., Flegal, K. M., Carroll, M. D., and Johnson, C. L. (2002). Prevalence and trends in overweight among US children and adolescents. *Journal of the American Medical Association* **288**, 1728–1732.
- Rosenbaum, P. R. and Rubin, D. B. (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika* **70**, 41–55.
- Rubin, D. B. (1987). *Multiple Imputation for Nonresponse in Surveys*. New York: John Wiley & Sons.
- Ruppert, D., Wand, M. P., and Carroll, R. J. (2003). *Semi-parametric regression*. Cambridge, U.K.: Cambridge University Press.
- Scott, A. and Smith, T. M. F. (1969). Estimation in multi-stage surveys. *Journal of the American Statistical Association* **64**, 830–840.
- Yuan, Y. and Little, R. J. A. (2007). Model-based estimates of the finite population mean for two-stage cluster samples with unit non-response. *Applied Statistics* **56**, 79–97.
- Zhang, G. and Little, R. J. (2005). Extensions of the penalized spline propensity prediction method of imputation. *American Statistical Association, Proceedings of the Joint Statistical Meetings*.
- Zheng, H. and Little, R. J. A. (2004). Penalized spline non-parametric mixed models for inference about a finite mean from two-stage samples. *Survey Methodology* **30**, 209–218.

Received June 2006. Revised January 2007.

Accepted February 2007.