Journal of Official Statistics, Vol. 24, No. 2, 2008, pp. 193–211

Model-Based Inference for Two-Stage Cluster Samples Subject to Nonignorable Item Nonresponse

Ying Yuan[1] *and Roderick J.A. Little*[2]

This article concerns item nonresponse adjustment for two-stage cluster samples, where nonresponse depends on covariates and underlying cluster characteristics, or on covariates and the missing outcome. In these circumstances, standard weighting and imputation adjustments are liable to be biased. To obtain consistent estimates, we propose extensions of the standard random-effects model for clustered data to model these two types of missing data mechanisms. These new methods are compared with existing approaches by simulation studies, and illustrated on data on household income from the Behavioral Risk Factor Surveillance System.

Key words: Random effects; multiple imputation; cluster-specific nonignorable; outcome-specific nonignorable.

1. Introduction

This article concerns item nonresponse adjustment for two-stage cluster samples under two types of nonignorable item nonresponse: (i) nonresponse depends on covariates and underlying cluster characteristics, and (ii) nonresponse depends on covariates and missing outcomes. These two types of nonignorable item nonresponse occur in practice, as the survey variable and nonresponse often share a common unobserved cause (Groves and Couper 1998). An example of this might be a two-stage cluster sample with counties as primary sampling units (PSU's) and households as secondary sampling units (SSU's). If nonresponse of households is deemed to be related to (unmeasured) county-level characteristics (e.g., geographic location, financial situation or goodness of administration) that are also associated with survey variables of interest, then the first type of nonresponse occurs. On the other hand, if the nonresponse of a household depends on the (unobserved or uncollected) household's characteristics such as lifestyle or genetic characteristics, which also affect the survey variables, then the resulted nonresponse is of the second type. The second type of nonignorable item nonresponse also arises when the cause of the nonresponse is the value of the survey variable itself. For example, households with high income may be less willing to divulge their income.

It is well-known that the standard weighting or regression methods generally lead to biased estimates when nonresponse is nonignorable (Groves and Couper 1998; Little and

[1] Department of Biostatistics, University of Texas, M.D. Anderson Cancer Center, Houston, TX, U.S.A. Email: yyuan@mdanderson.org
[2] Department of Biostatistics, The University of Michigan, Ann Arbor, MI, U.S.A. Email: rlittle@umich.edu
Acknowledgments: The authors are grateful to the editor and the referees for their constructive comments.

Rubin 2002). To deal with above two types of nonignorable nonresponse, Yuan and Little (2007a) propose model-based approaches to adjust unit nonresponse. This article extends their approaches to address item nonresponse.

Consider a finite population of size M consisting of N clusters with M_i elements in the ith cluster, let Y_{ij} denote the value of a survey outcome Y, and $X_{ij} = (X_{1ij}, \ldots, X_{Pij})$ denote values of P covariates $X = (X_1, \ldots, X_P)$ for unit j in the cluster i, for $i = 1, \ldots, N$; $j = 1, \ldots, M_i$. Let $T = \sum_{i=1}^{N}\sum_{j=1}^{M_i} Y_{ij}$ and $\bar{Y} = T/M$ denote the population total and mean, respectively. At the first stage, a sample of n of the N clusters (PSU's) is selected. At the second stage, m_i of the M_i units (SSU's) are selected in the ith 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 nonresponse depends on X and underlying cluster characteristics, or on X and the missing value of Y. The estimator of interest is the finite population mean $\bar{Y}$ or total T.

A common practical approach to estimating $\bar{Y}$ is to use the 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)

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

where π_{ij} is the selection probability of unit j in cluster i, determined by the survey design, and $\hat{y}_{ij}$ is the imputed value of y_{ij}. For this single imputation approach, the variance of $\hat{\bar{Y}}$ can be obtained by the adjusted jackknife variance estimation method (Rao and Shao 1992). Alternatively, multiple sets of draws can be imputed from the predictive distribution of y_{ij} under a model, and imputation uncertainty assessed using multiple imputation (MI). See, for example Rubin (1987) and Little and Rubin (2002).

The above method requires a model for predicting the missing values of Y. In particular for regression (REG) imputation, the value of y_{ij} for a nonrespondent is imputed by

$$\hat{y}_{ij} = \hat{\beta}_0 + \sum_{p=1}^{P}\hat{\beta}_P x_{pij} \tag{2}$$

where

$$(\hat{\beta}_0, \hat{\beta}_1, \ldots, \hat{\beta}_P)^T = \left(\sum_{i=1}^{n}\sum_{j=1}^{r_i}\pi_{ij}^{-1}x_{ij}^T x_{ij}\right)^{-1}\sum_{i=1}^{n}\sum_{j=1}^{r_i}\pi_{ij}^{-1}x_{ij}^T y_{ij}$$

with

$$x_{ij} = (1, x_{1ij}, x_{2ij}, \ldots, x_{Pij})$$

By substituting (2) into (1), we obtain the regression estimate $\hat{\bar{Y}}_{\text{REG}}$. This imputation is based on a regression model that ignores the clustering of the sample, and as a consequence $\hat{\bar{Y}}_{\text{REG}}$ is biased when nonresponse depends on underlying cluster characteristics. One way to correct the bias is to perform regression imputation separately for each cluster. In this case, the value of y_{ij} for a nonrespondent in the cluster i is imputed by

$$\hat{y}_{ij} = \hat{\beta}_{i0} + \sum_{p=1}^{P} \hat{\beta}_{ip} x_{pij} \tag{3}$$

where

$$\hat{\beta}_i = (\hat{\beta}_{i0}, \hat{\beta}_{i1}, \ldots, \hat{\beta}_{iP})^T = \left(\sum_{j=1}^{r_i} \pi_{ij}^{-1} x_{ij}^T x_{ij} \right)^{-1} \sum_{j=1}^{r_i} \pi_{ij}^{-1} x_{ij}^T y_{ij}$$

A drawback of this approach is that if r_i, the number of respondents in the cluster i, is small, $\hat{\beta}_i$ is not reliable. In the extreme case that r_i is less than the number of covariates P, $\hat{\beta}_i$ is not defined. In addition, if nonresponse directly depends on the missing value of Y, the resulting nonresponse is nonignorable and the regression imputation methods based on (2) or (3) are biased. In this article, we propose random-effects model-based approaches to address these problems.

Various approaches have been proposed in survey sampling literature to deal with nonignorable nonresponse. These approaches can be roughly divided into likelihood-based methods and weighting methods. For the likelihood-based approach, Little and Rubin (2002) further distinguish between selection models and pattern-mixture models, according to the way in which the joint distribution of the survey variables and missing data indicators is factorized. Greenlees, Reece, and Zieschang (1982) discuss a selection model for survey data with nonignorable nonresponse in which the probability of response depends on the missing value by a logistic regression model. Qin, Leung, and Shao (2002) study a semiparametric likelihood-based approach in which the response mechanism is modeled parametrically and the survey variable is modeled nonparametrically. Rubin and Zanutto (2002) propose a method using matched substitutes to adjust for nonignorable nonresponse through multiple imputations. Nandram and Choi (2002, 2005a) and Nandram, Cox, and Choi (2005b) propose hierarchical Bayesian selection models for categorical data with nonignorable nonresponse. Comparatively, weighting methods are more in line with conventional design-based approaches. The simple inverse-probability-weighting method weights the observed values of respondents by estimated response rates based on a response model. The resulting estimates are less efficient than those obtained from likelihood-based analysis and sensitive to the choice of weighting model (Clayton, Spiegelhalter, Dunn, and Pickels 1998). Robins, Rotnitzky, and Zhao (1995), Rotnitzky and Robins (1997), and Scharfstein, Rotnitzky, and Robins (1999) propose improved inverse-probability-weighted estimators, namely doubly robust estimators, to deal with ignorable and nonignorable nonresponse. Carpenter, Kenward, and Vansteelandt (2006) compare pros and cons of multiple imputation and inverse-probability-weighted methods.

The article is organized as follows. Section 2 presents our models for estimating the finite population mean when nonresponse depends on underlying cluster characteristics or missing outcome values. Section 3 describes estimation of the proposed models. Section 4 presents a simulation study comparing these methods with the existing methods described above. Section 5 illustrates the methods with the Behavioral Risk Factor Surveillance System data. Section 6 discusses our findings and conclusions.

2. Models

A natural way of addressing the lack of cluster effects in REG is to include covariates in the random-effects model (RE) first proposed for complete cluster samples by Scott and Smith (1969), as follows:

$$[y_{ij}|\alpha_i, \beta_0, \beta_p, \sigma^2] = N\left(\alpha_i + \beta_0 + \sum_{p=1}^{P} \beta_p x_{pij}, \sigma^2\right), \quad [\alpha_i|\tau^2] = N(0, \tau^2) \tag{4}$$

where $N(\cdot)$ denotes normal distribution. Under this model, inference for the finite population mean $\bar{Y}$ can be obtained by MI. Specifically, we first form K imputed datasets by filling in missing values of y with K independent draws from their posterior distribution based on Model (4). For the kth imputed dataset, $\bar{Y}$ is estimated by

$$\hat{\bar{Y}}_k = \sum_{i=1}^{n}\sum_{j=1}^{m_i} \frac{y_{ij}^{(k)}}{\pi_{ij}} \Big/ \sum_{i=1}^{n}\sum_{j=1}^{m_i} \frac{1}{\pi_{ij}}$$

where $y_{ij}^{(k)}$ is the observed or imputed value of y_{ij}. Then, a consistent estimator of $\bar{Y}$ is given by

$$\hat{\bar{Y}} = \frac{1}{K}\sum_{k=1}^{K} \hat{\bar{Y}}_k$$

and its variance is

$$\mathrm{Var}(\hat{\bar{Y}}) = \frac{1}{K}\sum_{k=1}^{K} V_k + \frac{1}{K-1}\sum_{k=1}^{K} (\hat{\bar{Y}}_k - \hat{\bar{Y}})^2$$

where V_k is the variance of $\hat{\bar{Y}}_k$ calculated for the kth imputed dataset.

It is well-known that RE yields a consistent estimator of $\bar{Y}$ when nonresponse is missing at random (MAR), providing the regression equation is correctly specified. Since 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 α_i, the missing-data mechanism is not MAR, since the random effects that characterize the clusters are not observed (Little and Rubin 2002, Example 6.24; Yuan and Little 2007a). A consequence of nonignorability is that RE leads to biased estimates. As in the latter article for the case of unit nonresponse with no covariates, we use the term cluster-specific nonignorable (CNI) nonresponse to describe the mechanism where the probability of response depends on underlying cluster effects α_i, and observed covariates, but not on survey outcomes within clusters. Furthermore, we use the term outcome-specific nonresponse (ONI) for the case where missingness depends directly on the value of the outcome variable Y.

We first consider nonresponse adjustment for CNI nonresponse. As noted above, RE leads to biased estimators in this case. To correct the bias, an informal approach is to modify the RE model to allow the cluster mean to depend on the estimated cluster response rate $\hat{\phi}_i = r_i/m_i$. Assuming for simplicity a linear relationship yields the approximate model

$$\begin{aligned} [y_{ij}|x_{ij}, \alpha_i, \delta, \beta_0, \beta_p, \sigma^2] &= N\left(\alpha_i + \delta\hat{\phi}_i + \beta_0 + \sum_{p=1}^{P} \beta_p x_{pij}, \sigma^2\right) \\ [\alpha_i|\tau^2] &= N(0, \tau^2) \end{aligned} \tag{5}$$

which is easily fitted by including the estimated response rate in each cluster as a covariate in the RE model. We label estimates for this model RERR, for the random-effects model with response rate as covariate. RERR is approximate in that the sampling error in estimating the response rate is not taken into account.

A more rigorous approach is to model the CNI missing data mechanism directly, yielding the following parametric cluster-specific nonignorable model (PCNI):

$$\begin{aligned} [y_{ij}|\alpha_i, \chi_i, \delta, \beta_0, \beta_p, \sigma^2] &= N\left(\alpha_i + \delta\chi_i + \beta_0 + \sum_{p=1}^{P} \beta_p x_{pij}, \sigma^2\right) \\ [z_{ij}|\chi_i, \gamma_0, \gamma_p] &= N\left(\chi_i + \gamma_0 + \sum_{p=1}^{P} \gamma_p x_{pij}, 1\right), \quad r_{ij} = \begin{cases} 1 & \text{if } z_{ij} > 0 \\ 0 & \text{if } z_{ij} < 0 \end{cases} \\ [\alpha_i|\tau^2] &= N(0, \tau^2), \quad [\chi_i|\omega^2] = N(0, \omega^2) \end{aligned} \tag{6}$$

where r_{ij} is the response indicator. $r_{ij} = 1$ for respondents and $r_{ij} = 0$ for nonrespondents. z_{ij} is a latent variable that determines the response status of the subject j in the cluster i. If the value of z_{ij} is larger than the threshold 0, the subject responds; otherwise, the subject does not respond. Random effects α_i and χ_i model within-cluster correlations. Note that 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. We propose to estimate $\bar{Y}$ by MI based on predictions of the missing values from the PCNI model.

The above models assume a CNI mechanism where r_{ij} and y_{ij} are unrelated within clusters after conditioning on covariates and χ_i. A missing data mechanism is outcome-specific nonignorable (ONI) if the missingness of y_{ij} depends directly on the value of y_{ij} and observed covariates. Such a mechanism is modeled by changing PCNI so that y_{ij} has a linear regression on z_{ij} rather than the random effect χ_i, that is:

$$\begin{aligned} [y_{ij}|\alpha_i, \delta, \beta_0, \beta_p, \sigma^2] &= N\left(\alpha_i + \delta z_{ij} + \beta_0 + \sum_{p=1}^{P} \beta_p x_{pij}, \sigma^2\right) \\ [z_{ij}|\chi_i, \gamma_0, \gamma_p] &= N\left(\chi_i + \gamma_0 + \sum_{p=1}^{P} \gamma_p x_{pij}, 1\right), \quad r_{ij} = \begin{cases} 1 & \text{if } z_{ij} > 0 \\ 0 & \text{if } z_{ij} < 0 \end{cases} \\ [\alpha_i|\tau^2] &= N(0, \tau^2), \quad [\chi_i|\omega^2] = N(0, \omega^2) \end{aligned} \tag{7}$$

We label this model PONI. Again, the finite population mean $\bar{Y}$ can be estimated by MI.

All models discussed so far assume constant regression slopes across clusters. A more flexible approach is to allow each cluster to have an individual slope, which is analogous to performing the regression imputation separately for each cluster as (3). However, again, one potential problem is that estimates of slopes are not reliable for clusters with sparse observations. To address this problem, we assume that these cluster-level individual slopes are exchangeable, in particular come from a normal distribution, to borrow strength across clusters. For example, RE with cluster-specific random slopes can be expressed as follows:

$$
\begin{aligned}
&[y_{ij}|x_{ij}, \alpha_i, b_{pi}, \beta_0, \beta_p, \sigma^2] = N\left(\alpha_i + \beta_0 + \sum_{p=1}^{P} \beta_p x_{pij} + \sum_{p=1}^{P} b_{pi} x_{pij}, \sigma^2\right) \\
&\quad [\alpha_i|\tau^2] = N(0, \tau^2), \quad [b_{pi}|\eta_p^2] = N(0, \eta_p^2)
\end{aligned}
\tag{8}
$$

where b_{pi} are cluster-specific random slopes. Of course, if some slopes are fixed across clusters a *priori*, we can simply set the corresponding b_{pi} as 0. We label this model RE2. The similar extension is readily applied to other models. In particular, for ONI nonresponse, the random slopes extension of PONI is given by

$$
\begin{aligned}
&[y_{ij}|x_{ij}, z_{ij}, \alpha_i, b_{pi}, \delta, \beta_0, \beta_p, \sigma^2] = N\left(\alpha_i + \delta z_{ij} + \beta_0 + \sum_{p=1}^{P} \beta_p x_{pij} + \sum_{p=1}^{P} b_{pi} x_{pij}, \sigma^2\right) \\
&[z_{ij}|x_{ij}, \chi_i, \gamma_0, \gamma_p] = N\left(\chi_i + \gamma_0 + \sum_{p=1}^{P} \gamma_p x_{pij}, 1\right) \quad r_{ij} = \begin{cases} 1 & \text{if } z_{ij} > 0 \\ 0 & \text{if } z_{ij} < 0 \end{cases} \\
&[\alpha_i|\tau^2] = N(0, \tau^2), \quad [\chi_i|\omega^2] = N(0, \omega^2), \quad [b_{pi}|\eta_p^2] = N(0, \eta_p^2)
\end{aligned}
\tag{9}
$$

3. Estimation of Models

A convenient approach to fitting the models described in Section 2 is to add noninformative or diffuse priors for the fixed parameters and simulate draws from the posterior distribution of the parameters. For recent reviews of the Bayesian approach to sample surveys, see, for example Little (2003, 2004). Estimates for RE are easily obtained by the Gibbs sampler discussed in Gelfand, Hills, Racine-Poon, and Smith (1990), and PCNI and PONI can also be fitted by the Gibbs sampler if we treat the latent variable Z as missing data. The convergence of the Gibbs chains is monitored by graphical inspection, and by the method of Gelman and Rubin (1992). We here use Model (9) as an example to illustrate this approach.

We first define priors for the fixed parameters of the form

$$
\begin{aligned}
&[\delta, \beta_0, \ldots, \beta_P, \gamma_0, \ldots, \gamma_P, \sigma^2, \tau^2, \omega^2] \\
&\quad \propto (\sigma^2)^{-(A+1)} e^{-B/\sigma^2} (\tau^2)^{-(A+1)} e^{-B/\tau^2} (\omega^2)^{-(A+1)} e^{-B/\omega^2} \\
&[\eta_p^2] \propto (\eta_p^2)^{-(A+1)} e^{-B/\eta_p^2}, \quad p = 1, \ldots, P
\end{aligned}
$$

where $A = B = 0.1$, a value small enough that the information in the data strongly dominates the information in the prior distribution. Let y_{obs}, y_{mis} denote values of the survey outcome Y for respondents and nonrespondents. The first step of the iteration is "data augmentation" (Tanner and Wong 1987), in which the missing y_{ij} and the latent variable z_{ij} are generated from their full conditional distributions. When $r_{ij} = 1$, that is, y_{ij} is observed, z_{ij} is drawn from the following left truncated normal distribution (TN):

$$[z_{ij}|y_{\text{obs}}, x_{ij}, \theta] = TN_{[z_{ij}>0]}(\mu_{z_{ij}}, \sigma^2_{z_{ij}})$$

where

$$\mu_{z_{ij}} = \frac{\delta\left(y_{ij} - \alpha_i - \beta_0 - \sum_{p=1}^{P} \beta_p x_{pij} - \sum_{p=1}^{P} b_{pi} x_{pij}\right) + \sigma^2\left(\chi_i + \gamma_0 + \sum_{p=1}^{P} \gamma_p x_{pij}\right)}{\sigma^2 + \delta^2}$$

$$\sigma^2_{z_{ij}} = \frac{\sigma^2}{\sigma^2 + \delta^2}, \qquad \theta = \{\alpha_i, \beta_0, \beta_p, b_{pi}, \delta, \chi_i, \gamma_0, \gamma_p, \sigma^2\}$$

When $r_{ij} = 0$, (y_{ij}, z_{ij}) are drawn from the following conditional distributions:

$$[z_{ij}|\chi_i, \gamma_0, \gamma_p] = TN_{[z_{ij}\le 0]}\left(\chi_i + \gamma_0 + \sum_{p=1}^{P} \gamma_p x_{pij}, 1\right)$$

$$[y_{ij}|z_{ij}, \alpha_i, b_{pi}, \delta, \beta_0, \beta_p, \sigma^2] = N\left(\alpha_i + \delta z_{ij} + \beta_0 + \sum_{p=1}^{P} \beta_p x_{pij} + \sum_{p=1}^{P} b_{pi} x_{pij}, \sigma^2\right)$$

Let 1_n denote a vector of 1 with length n, 0_n denote an $n \times n$ matrix with all elements of 0, and I_n denote an $n \times n$ identity matrix. We define the design matrix $X = (1, x_1, \ldots, x_P)$ where $x_p = (x_{p11}, \ldots, x_{p1m_1}, \ldots, x_{pn1}, \ldots, x_{pnm_n})^T$ for $p = 1, \ldots, P$, and block diagonal matrix $E = \text{blockdiag}(1_{m_1}, \ldots, 1_{m_i}, \ldots, 1_{m_n})$. Then, with augmented complete data, parameters are drawn as follows:

(1) draw $(\gamma_0, \gamma_1, \ldots, \gamma_P, \chi_1, \ldots, \chi_n)$ from

$$[\gamma_0, \gamma_1, \ldots, \gamma_P, \chi_1, \ldots, \chi_n|z, x, \omega^2] = N((C^T C + D)^{-1} C^T z, (C^T C + D)^{-1})$$

where $C = (X, E)$, $D = \text{blockdiag}(0_{P+1}, \omega^{-2} I_n)$ and $z = (z_{11}, \ldots, z_{1m_1}, \ldots, z_{n1}, \ldots, z_{nm_n})^T$

(2) draw ω^2 from

$$[\omega^2|\chi_1, \ldots, \chi_n] = IG\left(A + \frac{1}{2}n, B + \frac{1}{2}\sum_{i=1}^{n} \chi_i^2\right)$$

where IG (a,b) denotes an inverse-gamma distribution with a shape parameter a and a scale parameter b.

(3) letting $\beta = (\beta_0, \beta_1, \ldots, \beta_P, \delta)$ denote parameters of fixed effects, and $b = (\alpha_i, \ldots, \alpha_n, b_{11}, \ldots, b_{1n}, \ldots, b_{P1}, \ldots, b_{Pn})$ denote random effects, draw (β, b) from

$$[\beta, b|y, \tau^2, \eta_p^2] = N((C^{*T} C^* + \sigma^2 D^*)^{-1} C^{*T} y, \sigma^2 (C^{*T} C^* + \sigma^2 D^*)^{-1})$$

where

$$C^* = (X, z, E, X_1^*, \ldots, X_P^*)$$
$$D^* = \text{blockdiag}(0_{P+2}, \tau^{-2}I_n, \eta_1^{-2}I_n, \ldots, \eta_P^{-2}I_n)$$
$$X_p^* = \text{blockdiag}((x_{p11}, \ldots, x_{p1m_i})^T, \ldots, (x_{pn1}, \ldots, x_{pnm_n})^T), \quad p = 1, \ldots, P$$
$$y = (y_{11}, \ldots, y_{1m_1}, \ldots, y_{n1}, \ldots, y_{nm_n})^T$$

(4) draw τ^2 from

$$[\tau^2|\alpha_i] = IG\left(A + \frac{1}{2}n, B + \frac{1}{2}\sum_{i=1}^{n}\alpha_i^2\right)$$

(5) for $p = 1, \ldots, P$, draw η_p^2 from

$$[\eta_p^2|b_{pi}] = IG\left(A + \frac{1}{2}n, B + \frac{1}{2}\sum_{i=1}^{n}b_{pi}^2\right)$$

4. Simulation Study

4.1. Description of Simulation Study

We conducted a simulation study to compare the performance of various methods under different missing data mechanisms (missing completely at random (MCAR), missing at random (MAR), CNI, ONI, and mixture of CNI and ONI), mean models (linear fixed-slope, linear random-slope and cubic), and propensity models (linear and cubic). Thirty populations of $M = \sum_{i=1}^{N} M_1 = 40,492$ values of a variable Y were constructed in 200 clusters. Cluster sizes $\{M_i\}$ were randomly generated from a uniform distribution with a minimum size of 20 and a maximum size of 400. The populations were generated according to the model:

$$[y_{ij}|x_{ij}, z_{ij}, \alpha_i, \lambda, \delta, \sigma^2] = N(\alpha_i + \lambda\delta(z_{ij} - \chi_i) + \delta\chi_i + g_1(x_{1ij}, \ldots, x_{pij}), \sigma^2)$$

$$[z_{ij}|x_{ij}, \chi_i] = N(\chi_i + g_2(x_{1ij}, \ldots, x_{pij}), 1), \quad r_{ij} = \begin{cases} 1 & \text{if } z_{ij} > 0 \\ 0 & \text{if } z_{ij} < 0 \end{cases} \tag{10}$$

$$[\alpha_i|\tau^2] = N(0, \tau^2), \quad [\chi_i|\omega^2] = N(0, \omega^2)$$

where $g_1(\cdot)$ is a function that determines how the mean of Y depends on the covariates, and $g_2(\cdot)$ is a function that determines how the mean of the latent variable Z, and hence the probability of nonresponse, depends on the covariates. We varied $g_1(\cdot)$ and $g_2(\cdot)$ to simulate data with different mean models and propensity models. For convenience, we assumed one covariate x_1, which was sampled independently from N(2,1). The values of x_1 were observed for both respondents and nonrespondents. Specifically, we let $g_1(\cdot)$ take (1) a linear form of $g_1(x_{1ij}) = \beta_0 + \beta_1 x_{1ij}$ with either constant slopes $\beta_1 = 5$ or random slopes β_1 sampled from N(5,4), or (2) a cubic form of $g_1(x_{1ij}) = \beta_0^* + x_{1ij}^3$, where β_0 and β_0^* were chosen so that the superpopulation mean is 15. We let $g_2(\cdot)$ take (1) a linear form of $g_2(x_{1ij}) = \gamma_0 + \gamma_1 x_{1ij}$ with $\gamma_0 = 0.5$, $\gamma_1 = 0$ or $\gamma_0 = -2.2$, $\gamma_1 = 1$ to generate nonresponse independent or dependent of covariates; (2) a cubic form of $g_2(x_{1ij}) = 1.85 - 0.1x_{1ij}^3$. In all cases, the expected response rate was 60%. The parameter δ in (10) determines the degree of nonignorability of the missing-data mechanism. If $\delta = 0$, the missing data are ignorable. Specifically, if $\delta = 0$ and $g_2(x_{1ij})$ is a constant (e.g., $\gamma_1 = 0$), missing data are MCAR; if $\delta = 0$ and $g_2(x_{1ij})$ depends on x_{1ij} (e.g., $\gamma_1 \neq 0$), missing data are MAR. Nonzero values of δ specify nonignorable missing-data mechanisms, and larger values correspond to stronger degrees of nonignorability. We simulated populations with two values of δ, $\delta = 0$ and 5. The parameter λ determines the extent to which the missing-data mechanism depends on the cluster-level random effects χ_i and the value of y_{ij} itself. If $\lambda = 1$, the missingness of y_{ij} depends entirely on y_{ij} and x_{1ij} (e.g., ONI nonresponse); if $\lambda = 0$, the missingness of y_{ij} only depends on χ_i and x_{1ij} (e.g., CNI nonresponse); if $0 < \lambda < 1$, the missing mechanism lies somewhere between these two extremes (e.g., mixture of ONI and CNI nonresponse). We simulated data with three values of λ, $\lambda = 0$, 0.5 and 1. For the other parameters, we set $\sigma^2 = 1$, $\tau^2 = 1$ and $\omega^2 = 4$.

A two-stage equal selection probability design was applied to these populations. A first-stage sample of $n = 20$ PSUs (or clusters) was chosen by PPS, and a second-stage sample of $m = 10$ SSUs (or elements) was selected from each sampled PSU, yielding a total sample size of 200. The sampling scheme was repeated 500 times for each population. We chose $K = 30$ multiple imputations for model-based approaches. The estimates of $\bar{Y}$ from the following ten methods were computed:

(REG) regression estimator (2)
(REG2) regression estimator where regression imputation is performed separately for each cluster as (3)
(RE) random-effects model (4)
(RE2) random-effects model with random slopes (8)
(RERR) random-effects model with estimated response rate as covariate (5)
(RERR2) RERR with random slopes
(PCNI) parametric cluster-specific nonignorable model (6)
(PCNI2) parametric cluster-specific nonignorable model with random slopes
(PONI) parametric outcome-specific nonignorable model (7)
(PONI2) parametric outcome-specific nonignorable model with random slopes (9)

For the purpose of comparison, we also calculated the sample mean before deletion (BD) of the missing-data, and the widely used simple weighted estimator. The simple weighted estimator ignores covariates information and weights the value of respondents by the product of the sampling weight and the cluster-specific response weight, i.e., $w_{ij} = (\pi_{ij}\hat{\phi}_i)^{-1}$.

4.2. Simulation Results

Table 1 shows the empirical bias, root mean squared error (RMSE) and 95% confidence interval coverage rate of each method over the 500 samples for the population with the linear (fixed-slopes) mean model and the linear propensity model, under various missing-data mechanisms. Table 2 shows the same summary statistics for the population with

Table 1. Empirical bias (× 100), RMSE (× 100) and coverage rate (× 100) of 95% confidence intervals for the population with constant slopes across clusters and the linear propensity model

Missing mechanism		Methods											
		BD	WT	REG	REG2	RE	RE2	RERR	RERR2	PCNI	PCNI2	PONI	PONI2
MCAR	Bias	−3	−2	2	−1	−3	−4	−9	−15	−9	−9	−15	−13
	RMSE	41	66	44	52	44	44	47	59	48	48	59	60
	Coverage	95.0	93.8	94.8	95.4	95.6	95.2	94.6	94.4	94.8	94.6	94.8	93.8
MAR	Bias	−3	209	1	−1	−3	2	−5	−6	−5	−5	−7	−6
	RMSE	41	220	44	76	42	42	43	49	44	44	51	51
	Coverage	95.0	9.2	94.8	95.6	95.6	95.4	96.4	95.8	96.2	96.4	95.6	96.6
CNI	Bias	−6	367	512	246	139	138	14	13	−13	−16	99	98
	RMSE	214	395	537	311	231	229	201	205	229	240	214	216
	Coverage	95.4	37.8	18.2	73.4	90.4	90.0	93.8	93.8	95.2	94.6	93.0	92.4
CNI & ONI	Bias	−7	579	592	385	299	319	163	172	145	154	47	57
	RMSE	219	592	610	430	338	353	238	242	242	249	214	212
	Coverage	95.2	0.0	3.4	36.8	63.0	56.8	86.8	85.6	88.4	87.8	94.2	94.0
ONI	Bias	−8	791	671	529	466	512	308	322	295	313	9	15
	RMSE	226	799	686	567	486	529	345	356	342	355	241	238
	Coverage	95.2	0.0	0.2	15.0	12.8	3.8	56.0	51.6	62.6	58.0	95.0	95.2

Table 2. Empirical bias (× 100), RMSE (× 100) and coverage rate (× 100) of 95% confidence intervals for the population with random slopes across clusters and the linear propensity model

Missing mechanism		Methods											
		BD	WT	REG	REG2	RE	RE2	RERR	RERR2	PCNI	PCNI2	PONI	PONI2
MCAR	Bias	−5	13	−17	3	4	−10	22	−27	30	8	38	3
	RMSE	90	114	112	108	101	102	114	116	121	105	206	121
	Coverage	94.6	94.6	92.2	93.2	93.2	94.0	94.2	93.8	94.8	93.8	93.6	93.2
MAR	Bias	−5	248	−12	19	12	−1	23	−15	27	3	45	2
	RMSE	90	278	106	127	97	96	114	107	119	99	198	108
	Coverage	94.6	53.8	93.0	94.4	96.6	93.4	96.0	93.8	95.8	94.8	93.8	94.6
CNI	Bias	−18	401	566	289	167	150	20	−10	−16	−19	60	92
	RMSE	242	448	604	366	282	266	235	235	262	256	266	244
	Coverage	93.8	50.8	29.0	77.0	87.6	88.8	93.8	93.4	94.4	94.2	91.2	91.8
CNI & ONI	Bias	−19	631	647	426	322	366	182	187	148	173	16	57
	RMSE	247	655	677	478	379	411	277	278	272	274	248	241
	Coverage	94.2	4.8	11.2	44.4	70.6	57.0	86.4	84.8	88.8	88.4	94.0	92.8
ONI	Bias	20	861	728	570	487	545	341	360	306	335	−15	1
	RMSE	254	877	753	612	518	570	392	406	368	387	279	263
	Coverage	94.8	0	1.4	18.8	32.4	12.0	64.4	57.4	73.2	68.2	95.0	94.8

the linear (random-slopes) mean model and the linear propensity model. To save space, we omit detailed results for the populations when the mean model or/and the propensity model are misspecified, but show plots of empirical bias for the cubic mean and linear propensity in Figure 1, the linear fixed-slope mean and cubic propensity in Figure 2, and the cubic mean and cubic propensity in Figure 3.

4.2.1. MCAR Nonresponse ($\delta = \lambda = \gamma_1 = 0$)

When nonresponse is MCAR, all methods are valid, yielding unbiased estimates and reasonable coverage rates, given that both the mean model and the propensity model are linear. RE is the most efficient, due to borrowing strength across clusters and the parsimonious model specification.

4.2.2. MAR Nonresponse ($\delta = \lambda = 0$)

The missingness of y_{ij} only depends on the observed x_{ij}, and the missing-data mechanism is MAR. When both the mean model and the propensity model are linear, all methods except WT yield unbiased estimates and close to nominal coverage rates. WT is subject to bias because it does not account for covariates. When slopes are constant across clusters in the population, as shown in Table 1, REG2 is less efficient than REG, but models with random slopes, such as RE2, RERR2, PCNI2, and PONI2, do not suffer a substantial efficiency loss compared to the counterparts with fixed slopes, reflecting the advantage of borrowing strength across clusters. When slopes vary across clusters, as shown in Table 2, methods assuming constant slopes, like RE, RERR, PCNI and PONI, still yield unbiased estimates, but have a larger RMSE than the counterparts with random slopes. Figure 2 suggests that if the mean model is correctly specified, the misspecification of the propensity model has little effect on the performance of any of the methods when nonresponse is MAR. In contrast, all the methods are sensitive to the assumption that

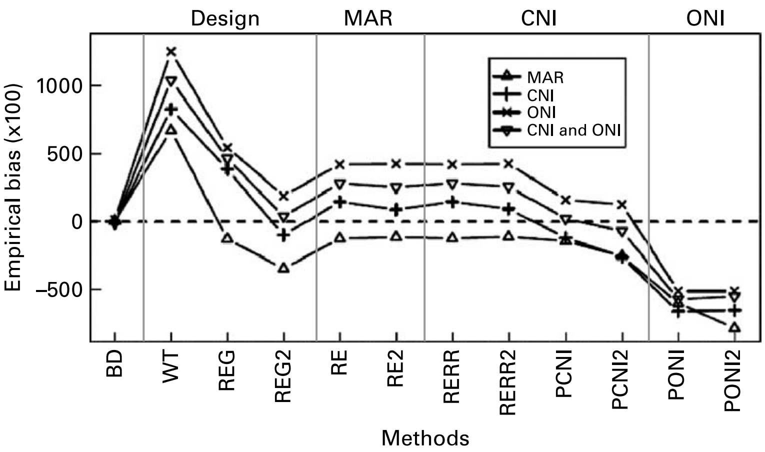

Fig. 1. Empirical bias (× 100) for eleven methods under various missing mechanisms for the population with the cubic mean structure and the linear propensity model. The dashed line denotes the value of 0. As shown by labels on the top of the panel, these methods are grouped into four categories: Design-based methods, methods assuming MAR, methods assuming CNI and methods assuming ONI

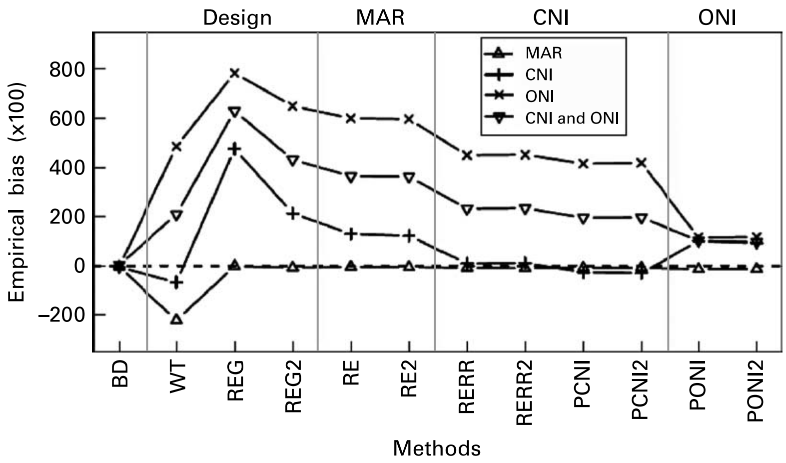

Fig. 2. Empirical bias (× 100) for eleven methods under various missing mechanisms for the population with the linear mean structure and the cubic propensity model. The dashed line denotes the value of 0. As shown by labels on the top of the panel, these methods are grouped into four categories: Design-based methods, methods assuming MAR, methods assuming CNI and methods assuming ONI

$E(y_{ij}|x_{ij})$ is a linear function of x_{ij}, and lead to biased estimates when the assumption does not hold (Figure 1). As expected, when both the mean model and the propensity model are misspecified as in Figure 3, all the methods lead to biased estimates.

4.2.3. CNI Nonresponse ($\delta = 5, \lambda = 0$)

In this case, nonresponse is associated with unobserved cluster characteristics. As displayed by Table 1, WT, REG, RE, and RE2 yield biased estimates due to the lack of

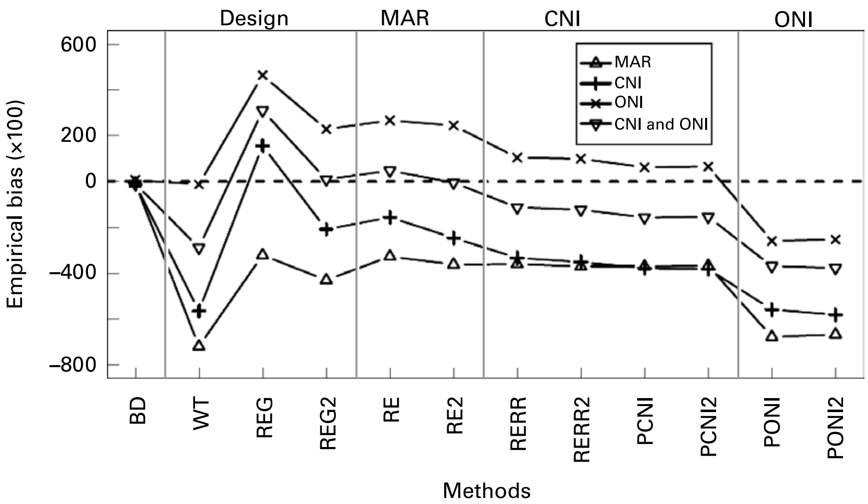

Fig. 3. Empirical bias (× 100) for eleven methods under various missing mechanisms for the population with the cubic structure and the cubic propensity model. The dashed line denotes the value of 0. As shown by labels on the top of the panel, these methods are grouped into four categories: Design-based methods, methods assuming MAR, methods assuming CNI and methods assuming ONI

CNI nonresponse adjustment. REG2 is biased because the model drops clusters with one or zero observation. Nevertheless, the bias is smaller than that of REG. RERR corrects the bias of RE, but the confidence intervals slightly undercover the population mean since this method ignores the uncertainty of estimated response rates. PCNI and PCNI2 take account of CNI nonresponse and yield unbiased estimates and sound coverage rates. PCNI has a slightly smaller RMSE when slopes are constant across clusters, while PCNI2 has a slightly smaller RMSE as slopes vary across clusters. Figure 2 suggests that PCNI and PCNI2 are robust to the misspecification of the propensity model, but sensitive to the misspecification of the mean model as illustrated by Figure 1. They lead to biased estimates when y_{ij} is not linearly associated with x_{ij}. PONI and PONI2 are slightly biased due to the misspecification of nonignorable missing-data mechanism.

4.2.4. ONI Nonresponse ($\delta = 5, \lambda = 1$)

In this situation, nonresponse is related to individual outcomes within clusters, and PONI and PONI2 are the only methods that correctly specify the ONI mechanism. As a result, PONI and PONI2 are the best methods in terms of RMSE and coverage, and the other methods all perform poorly. Unfortunately, PONI and PONI2 are sensitive to misspecification of the mean model (Figure 1), the propensity model (Figure 2) or both (Figure 3).

4.2.5. Mixed CNI and ONI Nonresponse ($\delta = 5, \lambda = 0.5$)

This nonignorable missing data mechanism is a mixture of CNI and ONI nonresponse. None of the methods are satisfactory in terms of bias, RMSE and coverage in this case. Thus there is no single method that dominates the others consistently over the simulation conditions.

5. Application

The Behavioral Risk Factor Surveillance System (BRFSS) is a collaborative project of the Centers for Disease Control and Prevention (CDC) and U.S. states and territories. The objective of the BRFSS is to collect uniform, state-specific data on preventive health practices and risk behaviors that are linked to chronic diseases, injuries, and preventable infectious diseases in the adult population. Most states (including Tennessee) in BRFSS used a disproportionate stratified sample design. In this design, telephone numbers are divided into high-density or medium-density strata, and are sampled separately to obtain a probability sample of all households with telephones.

We consider here estimating mean annual household income (from all sources) in Tennessee based on BRFSS public-use data collected in 2003. In this data set, household income is the variable with the largest percentage of missing values, about 22% nonresponse. Other good predictors of income, namely age, education, gender, race, employment status and number of adults in household, have nonresponse rates of less than 1%, and are used as covariates to predict the missing values of income. Linear and quadratic effects of age are included, and all other covariates are treated as categorical. Income is in principle continuous, but is here reported in eight categories because of confidentiality concerns. We treated it as continuous with values given by the medians in each category. We applied a cube root transformation to improve normality, as in Schenker et al. (2005) and Paulin and Sweet (1996). Although Tennessee does not use a two-stage cluster sampling design, sample subjects are clustered naturally by counties, and these are modeled by random effects in our methods.

As an empirical analysis of the missing-data mechanism of nonresponse, we compared distributions of covariates between respondents and nonrespondents. The analysis shows that characteristics of nonrespondents are significantly different from those of respondents. The mean ages of respondents and nonrespondents are 53.4 and 46.8, respectively, corresponding to a p-value of less than 0.0001. Respondents and nonrespondents are also significantly different in terms of gender, education, and employment status (the χ^2 test yielded p-values of 0.008, 0.004, and <0.0001, respectively). For example, 37.7% of respondents versus 31.6% of nonrespondents are men, and 25.2% respondents versus 19.3% of nonrespondents are college graduates. These results suggest that the nonresponse in BRFSS is not MCAR. To further assess the missing-data mechanism, we fit the RE model and plotted the average of the resulting regression residuals within counties against the response rate of the counties. In this plot, a systematic pattern suggests that nonresponse may be related to cluster-specific characteristics, that is, CNI nonresponse. For this particular data, we did not observe any systematic pattern, providing certain empirical evidence that the nonresponse here may not be CNI. From a statistical point of view, however, there is no information to distinguish between MAR, CNI, and ONI. In this case, other published studies may provide some prior information about the missing-data mechanism. Greenlees, Reece, and Zieschang (1982), Lillard, Smith, and Welch (1986), and Neukirch (2002) reported that individuals with higher income tend to have smaller probabilities of response, i.e., the nonresponse is nonignorable.

Table 3 shows estimates of the finite population mean of household income in Tennessee and associated standard errors under different models. The weighting estimate WT is not recommended, since the nonresponse is not MCAR. Methods other than WT and the outcome-specific nonignorable models PONI and PONI2 yielded similar estimates of the average income of about \$41,000. The estimates of PONI and PONI2 are slightly larger than those of other models, yielding an average income of about \$42,800. The random-slopes models produce results very similar to those produced by their fixed-slopes counterparts, suggesting homogeneous slopes across clusters. Since the observed data do not contain any empirical evidence to distinguish the ignorable and nonignorable missing-data mechanism, given that some prior studies (Greenlees, Reece, and Zieschang 1982; Lillard, Smith, and Welch 1986; Neukirch 2002) suggest that income nonresponse tends to be ONI, it might be reasonable to assume the nonresponse in BRFSS as ONI and report the estimate of the average household income in Tennessee as \$42,800. Alternatively, 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 set, we might report that the average household income in Tennessee is in the range of \$41,000 to \$42,800.

In this example, the estimates from ONI models are not substantially different from the estimates based on other ignorable models. This may be because the nonignorability of nonresponse is considerably weakened by conditioning on variables that are good predictors of income.

Table 3. Estimates of the finite population mean of household income in Tennessee by different methods. $\hat{\bar{Y}}$ is the estimate of the household income, and SE($\hat{\bar{Y}}$) is the estimate of the associated standard error

Estimates	Methods										
	WT	REG	REG2	RE	RE2	RERR	RERR2	PCNI	PCNI2	PONI	PONI2
$\hat{\bar{Y}}$	42,151	40,846	40,962	41,013	41,017	41,022	41,026	40,967	41,021	42,796	42,833
SE($\hat{\bar{Y}}$)	645	633	673	586	574	562	566	580	578	687	672

6. Conclusion

We have described some new methods for handling item nonresponse adjustment for two-stage cluster samples that take into account covariate and cluster information, and allow for nonignorable nonresponse depending on the cluster, as in the PCNI and PCNI2 models, or the outcome itself, as in the PONI and PONI2 models. For cluster-specific nonignorable nonresponse, misspecification of the mean model has much more effect on the performance of PCNI and PCNI2 than the misspecification of the propensity model; also PONI and PONI2 were generally biased, due to the misspecification of the nonignorable missing-data mechanism. When nonresponse was ONI, all methods except PONI and PONI2 were biased and had a poor coverage rate, but PONI and PONI2 were sensitive to misspecification of the mean model or the propensity model. The methods based on models with random slopes (PCNI2, PONI2) did not show much advantage over methods based on models with fixed slopes (PCNI, PONI) in the simulation study. However, the random slopes models may be more beneficial if we are also interested in estimating the mean within clusters, as in small area estimation.

A natural question given these findings is whether we can determine whether nonresponse is CNI or ONI, and which model should be used. Unfortunately, these two types of nonignorable nonresponse are not easily distinguished on the bases of the observed data. If auxiliary variables for nonrespondents and respondents are available, for example from census data, then we can compare the residual distribution of auxiliary variables, obtained by regressing on appropriate covariates, for 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 PCNI2; otherwise, we may consider PONI or PONI2. If we do not have external information to distinguish between alternative nonignorable missing mechanisms, it may be more appropriate to apply more than one method and compare the results.

Cognitive and social-psychological theory of survey participation provides a framework for understanding alternative nonignorable nonresponse mechanisms (Groves and Couper 1998; Tourangeau, Rips, and Rasinski 2000). Nonignorable nonresponse is most likely in self-administered questionnaire surveys, since the householder has the opportunity to review the entire questionnaire before making the participation decision. If the underlying cause of refusal to participate is the psychological threat posed by the topic in the survey, then the resulting nonresponse is most likely to be ONI since the sample subjects' attributes on the survey variables are determining the probability of their participating. In contrast, nonresponse in many interview-administrated surveys may be less directly related to the survey variables because the respondent has only a general idea about the topic of the survey.

One limitation of the proposed methods is that the multiple imputations are potentially sensitive to violations of the model assumptions. To relax the parametric assumptions about the mean structure of y_{ij} as a function of covariates, one might model $E(y_{ij}|x_{ij})$ nonparametrically, for example using a spline on the estimated propensity to respond (Yuan and Little 2007b). The normality assumption may be improved by a judicious choice of transformation, as in the example. Alternatively, other choices of the error distribution may be chosen; for example with a binary outcome, the proposed mean structure could be incorporated with a binomial error structure with a logistic function of the covariates.

This article focuses on item nonresponse and assumes the covariates are completely observed. However, in practice, covariates are often partially observed. In this case, if nonresponse follows a monotone pattern of missing values, the regression models discussed here could be applied sequentially, filling in the missing values from the most observed to the least observed variables, and conditioning on previously imputed variables in the sequence. For nonresponse with a general pattern of missing values, the sequential imputation approach (Raghunathan et al. 2001) could be used in conjunction with the proposed models.

7. References

Carpenter, J.R., Kenward, M.G., and Vansteelandt, S. (2006). A Comparison of Multiple Imputation and Double Robust Estimation for Analyses with Missing Data. Journal of the Royal Statistical Society, Series A, 169, 571–584.

Clayton, D., Spiegelhalter, D., Dunn, G., and Pickles, A. (1998). Analysis of Longitudinal Binary Data from Multiphase Sampling. Journal of the Royal Statistical Society, Series B, 60, 71–87.

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.B. (1992). Inference from Iterative Simulation Using Multiple Sequences. Statistical Science, 7, 457–511.

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, 82, 251–261.

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

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.

Lillard, L., Smith, J.P., and Welch, F. (1986). What Do We Really Know About Wages? The Importance of Nonreporting and Census Imputation. Journal of Political Economy, 94, 489–506.

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). New York: Wiley, 49–57.

Little, R.J.A. (2004). To Model or Not to Model? Competing Modes of Inference for Finite Population Sampling. Journal of the American Statistical Association, 99, 546–556.

Little, R.J.A. and Rubin, D.B. (2002). Statistical Analysis with Missing Data (2nd Edition). New York: Wiley.

Nandram, B. and Choi, J.W. (2002). Hierarchical Bayesian Nonresponse Models for Binary Data from Small Areas with Uncertainty about Ignorability. Journal of the American Statistical Association, 97, 381–388.

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

Nandram, B., Cox, L.H., and Choi, J.W. (2005b). Bayesian Analysis of Nonignorable Missing Categorical Data: An Application to Bone Mineral Density and Family Income. Survey Methodology, 31, 213–225.

Neukirch, T. (2002). Nonignorable Attrition and Selectivity Biases in the Finnish Subsample of the ECHP: An Empirical Study Using Additional Register Information, Working Paper 5, CHINTEX.

Paulin, G.D. and Sweet, E.M. (1996). Modeling Income in the U.S. Consumer Expenditure Survey. Journal of Official Statistics, 12, 403–419.

Qin, J., Leung, D., and Shao, J. (2002). Estimation with Survey Data under Nonignorable Nonresponse or Informative Sampling. Journal of the American Statistical Association, 97, 193–200.

Raghunathan, T.E., Lepkowski, J.M., Van Hoewyk, J., and Solenberger, P. (2001). A Multivariate Technique for Multiply Imputing Missing Values Using a Sequence of Regression Models. Survey Methodology, 27, 85–95.

Rao, J.N.K. and Shao, J. (1992). Jackknife Variance Estimation with Survey Data under Hot Deck Imputation. Biometrika, 79, 811–822.

Robins, J.M., Rotnitzky, A., and Zhao, L.P. (1995). Analysis of Semiparametric Regression Models for Repeated Outcomes in the Presence of Missing Data. Journal of the American Statistical Association, 90, 106–129.

Rotnitzky, A. and Robins, J.M. (1997). Analysis of Semi-Parametric Regression Models with Nonignorable Nonresponse. Statistics in Medicine, 16, 81–102.

Rubin, D.B. (1987). Multiple Imputation for Nonresponse in Surveys. New York: Wiley.

Rubin, D.B. and Zanutto, E. (2002). Using Matched Substitutes to Adjust for Nonignorable Nonresponse through Multiple Imputations. In Survey Nonresponse, R. Groves, D.A. Dillman, R.J.A. Little, and J. Eltinge (eds). New York: Wiley, 389–402.

Scharfstein, D.O., Rotnitzky, A., and Robins, J.M. (1999). Adjusting for Nonignorable Dropout Using Semiparametric Nonresponse Models (with Comments). Journal of the American Statistical Association, 94, 1096–1146.

Schenker, N., Raghunathan, T.E., Chiu, P., Makuc, D.M., Zhang, G., and Cohen, A.J. (2005). Multiple Imputation of Missing Income Data in the National Health Interview Survey. http://www.cdc.gov/nchs/data/nhis/tecdoc1.pdf

Scott, A. and Smith, T.M.F. (1969). Estimation in Multi-Stage Surveys. Journal of the American Statistical Association, 64, 830–840.

Tanner, M.A. and Wong, W.H. (1987). The Calculation of Posterior Distributions by Data Augmentation. Journal of the American Statistical Association, 82, 528–550.

Tourangeau, R., Rips, L.J., and Rasinski, K. (2000). The Psychology of Survey Response. New York: Cambridge University Press.

Yuan, Y. and Little, R.J.A. (2007a). Model-Based Estimates of the Finite Population Mean for Two-Stage Cluster Samples with Unit Nonresponse. Applied Statistics, 56, 79–97.

Yuan, Y. and Little, R.J.A. (2007b). Parametric and Semiparametric Model-Based Estimates of the Finite Population Mean for Two-Stage Cluster Samples with Item Nonresponse. Biometrics, 63, 1172–1180.

Received September 2006
Revised December 2007