Journal of Official Statistics, Vol. 27, No. 2, 2011, pp. 255–277

A Pseudo-GEE Approach to Analyzing Longitudinal Surveys under Imputation for Missing Responses

Iván A. Carrillo[1], *Jiahua Chen*[2], *and Changbao Wu*[3]

This article presents a pseudo-GEE approach to the analysis of longitudinal surveys when the response variable contains missing values. A cycle-specific marginal hot-deck imputation method is proposed to fill in the missing responses and a pseudo-GEE method is applied to the imputed data set. Consistency of the resulting pseudo-GEE estimators is established under a joint randomization framework. Linearization variance estimators are also developed for the pseudo-GEE estimators under the assumption that the finite population sampling fraction is small or negligible. Finite sample performances of the proposed estimators are investigated through an extensive simulation study using data from the National Longitudinal Survey of Children and Youth.

Key words: Complex sampling design; consistency; generalized estimating equations; joint randomization; hot-deck imputation; superpopulation model; variance estimation.

1. Introduction

Longitudinal surveys are an important tool for population studies where the primary interest is to examine population changes over time at the individual level. The power of the added dimension over time allows the separation of age and cohort effects (Diggle et al. 2002; Hedeker & Gibbons 2006) or the effect of treatments and population interventions from other potential confounders. A major theme in the design and analysis of longitudinal studies is to establish certain association or causation between a response variable and a group of predictors and to further identify important factors relating to the response variable. The generalized estimating equation (GEE) method, first proposed by Liang & Zeger (1986) for nonsurvey data, is a popular statistical inference tool for longitudinal studies. The approach is semi-parametric, involving assumptions similar to the generalized linear models but allows the use of working variance-covariance matrix for repeated measurements within the same subject, which does not necessarily coincide with the true correlation structure. The method has been widely used by social scientists and health researchers for analyzing longitudinal data from various population studies.

[1] National Institute of Statistical Sciences, 19 T.W. Alexander Drive, P.O. Box 14006 Research Triangle Park, NC 27709, U.S.A. Email: ivan@niss.org
[2] Department of Statistics, University of British Columbia, 333-6356 Agricultural Road, Vancouver, BC V6T 1Z2, Canada. Email: jhchen@stat.ubc.ca
[3] Department of Statistics and Actuarial Science, University of Waterloo, 200 University Avenue West, Waterloo, ON N2L 3G1, Canada. Email: cbwu@uwaterloo.ca
Acknowledgments: This research was supported by grants from the Natural Sciences and Engineering Research Council of Canada and an internship grant to the first author by Mathematics of Information Technology and Complex Systems (MITACS) and National Program for Complex Data Structures (NPCDS).

In a recent paper Carrillo et al. (2010) presented a pseudo-GEE approach to complex longitudinal surveys by incorporating the survey weights into the estimating equations. While the use of survey weights under the estimating equation approach has also been examined by several other authors, including Godambe and Thompson (1986), Binder and Patak (1994) and Godambe (1995), Carrillo et al. (2010) were the first to rigorously establish the consistency of the resulting pseudo-GEE estimator under a joint randomization framework. Linearization variance estimators were also developed. Rubin-Bleuer and Schiopu Kratina (2005) discussed the consistency of the pseudo-GEE estimator under different assumptions.

One of the major problems of longitudinal studies is missing values. This is a common issue in large scale cross-sectional surveys (see, for example, Groves et al. 2002), and the problem intensifies for longitudinal surveys. As Song (2007) puts it, "It is more difficult to deal with missing data in longitudinal studies. This is because missing data patterns appear much more sophisticated than those in cross-sectional studies." The usual kinds of missing values in cross-sectional studies are unit nonresponse and item nonresponse. The latter includes missingness on the main study variable or covariates. For longitudinal studies, in addition to unit nonresponse, missing patterns include attrition (i.e., drop-out completely after a certain cycle), intermittent missingness (i.e., persons do not respond at some cycles but remain in the study for other cycles), missing values for some variables at any particular cycle, and combinations of these missing types.

The following example illustrates some of the missing data scenarios described above. The National Longitudinal Survey of Children and Youth (NLSCY) is designed by Human Resources Development Canada to measure child development and well-being. Data from seven biennial cycles of the survey conducted from 1994 to 2007 are now available through Statistics Canada's Research Data Centers. One of the main objectives of the survey is to study the development of children's behaviour problems as they grow and examine factors that contribute to changes. A very important variable included in NLSCY data sets is PAS, Physical Aggression Score, which is derived from six to eight questions (depending on the age group) included in the survey. Earlier studies (Thomas 2004; Carrillo et al. 2005) found some significant factors contributing to change in aggressive behaviours as children grow. In a recent study, Carrillo-García (2006) examined the following nine covariates as potentially significant in explaining PAS: Age, Age2 (the square of Age), Depression Score of the PMK (person most knowledgeable-about the child), Punitive Parenting Status, Region, Gender, Family Status, Household Income Status, and Hours in Daycare. See Carrillo-García (2006) for further detail.

Table 1 shows the break-down of all respondents at each cycle with respect to the missingness of the main study variable (PAS) and the nine covariates for the first four

Table 1. Missing frequencies of PAS and 9 covariates in the NLSCY dataset

	PAS observed				PAS missing			
Cycle	1	2	3	4	1	2	3	4
All covariates observed	5,263	4,736	4,182	3,905	68	32	14	31
At least one covariate missing	124	278	416	233	115	40	331	254

cycles. Children who missed an entire cycle are not included in the table. For instance, there were 5,570 children surveyed at the initial cycle one, and they were classified into four groups: (i) 5,263 responded to PAS and all nine covariates; (ii) 124 responded to PAS but not to some covariates; (iii) 68 responded to all covariates but not to PAS; and (iv) 115 responded to neither PAS nor all covariates. Similarly, there were 4,423 children surveyed at cycle four, and the break-down numbers into the four groups are 3,905, 233, 31 and 254, respectively. Due to intermittent missingness, i.e., children missed one or more cycles of the survey but participated in other cycles, the total number of children who participated throughout all four cycles is 4,165. However, there are only 3,049 so-called "completers" who not only participated throughout all four cycles but also provided complete responses to all questions at each cycle. The naïve or "complete case analysis" approach has often been used, where only the data from "completers" are used for analysis. This amounts to deleting all individuals who either missed a cycle or failed to respond to certain questions. It is apparent that complete case analysis is inefficient and it is valid only if the data are missing completely at random.

This article extends the pseudo-GEE method presented in Carrillo et al. (2010) for the analysis of longitudinal surveys to situations where the response variable contains missing values. In Section 2, we describe a cycle-specific random hot-deck imputation procedure which makes the pseudo-GEE method a valid inference tool for analyzing longitudinal surveys under the proposed procedure. Both weighted and unweighted random hot-deck imputation methods are considered. Consistency of the pseudo-GEE estimator using the imputed data set is established in Section 3 under a joint randomization framework. Linearization variance estimators are developed in Section 4. Results from an extensive simulation study on the finite sample performances of the pseudo-GEE estimator and the proposed variance estimators are reported in Section 5, using simulation models built on the basis of the first four cycles of NLSCY data sets. Some concluding remarks are provided in Section 6. Proofs of major results are given in the Appendix.

2. Cycle-Specific Marginal Hot-deck Imputation

Let $U = \{1, 2, \cdots, N\}$ be the set of labels for the N subjects in the finite population. Let $(Y_{ij}; X_{ij1}, \cdots, X_{ijp})'$ be values of the response variable Y and the vector of p covariates $(X_1, \cdots, X_p)'$ for the ith subject at the time of the jth cycle of the survey, $j = 1, \cdots, T_i$. The T_i can be different for different subjects but in many studies $T_i = T$ is common for all subjects. This is typically the case for large-scale surveys and will be assumed for the rest of the article. Let s be the set of n subjects selected from the finite population by a complex sampling design; let $w_i = 1/P(i \in s)$ be the basic design weights; let $\{(Y_{ij}; X_{ij1}, \cdots, X_{ijp}), j = 1, \cdots, T, i \in s\}$ be the data set from the longitudinal survey.

We consider cases where values of covariates $(X_{ij1}, \cdots, X_{ijp})'$ are observed for all $i \in s$ at all cycles ($j = 1, \cdots, T$) but the response variable Y_{ij} is subject to missingness. This is motivated by the fact that under longitudinal surveys, values of covariates are less likely to be missing or can be filled in at a later stage without error. For time-independent covariates such as gender the likelihood of missing at all cycles is very small. For time-varying covariates such as age it is usually easy to fill up the gap when the variables are observed for some cycles but missing for some other cycles. At any particular cycle j, we assume

that Y_{ij} is missing at random, i.e., $P(R_{ij} = 1|Y_{ij}; X_{ij1}, \cdots, X_{ijp}) = P(R_{ij} = 1|X_{ij1}, \cdots, X_{ijp})$, where $R_{ij} = 1$ if Y_{ij} is observed and $R_{ij} = 0$ if Y_{ij} is missing.

It is common practice at large survey organizations such as Statistics Canada to create and release public use data files with missing values handled by imputation. The imputed data sets provide a common frame for studies with different objectives and can be analyzed by standard softwares. For longitudinal surveys, handling missing values by imputation makes it possible to use a single set of survey weights for different analyses. This is important since different types of weights are available for longitudinal survey data, including longitudinal weights up to a certain cycle and cross-sectional weights for each cycle, and the decision on which set of weights to use for a particular analysis is not straightforward. Another advantage of using a common imputed data file is that different analyses can be compared with each other and some internal consistency can be preserved. The most crucial part of any imputation procedure for longitudinal surveys, however, is to facilitate related statistical analyses based on the imputed data sets and to obtain valid and more efficient statistical inferences.

Hot-deck imputation is a procedure in which missing items are replaced by values from respondents. Ford (1983) and Sande (1983) contain detailed description of hot-deck imputation procedures. For longitudinal surveys with missing responses, we propose to use a cycle-specific marginal hot-deck imputation procedure. The method, combined with the pseudo-GEE approach presented in the next section, provides a satisfactory solution to the problem of missing values in response.

Our major assumption for the cycle-specific marginal imputation procedure is that all covariates $X_1, \cdots, X_p$ are either categorical or ordinal. We will discuss scenarios when one or more covariates are continuous in Section 6. Let c_k be the number of categories or possible values taken by X_k; let $G = c_1 \times \cdots \times c_p$. For a specific cycle j, the proposed hot-deck imputation procedure is as follows:

i) Divide the overall sample s into G nonoverlapping subsamples such that $s = \cup_{g=1}^{G} s_{jg}$ according to the cross-classified imputation cells by the p covariates. Let n_{jg} be the size of s_{jg}. Note that $n = \sum_{g=1}^{G} n_{jg}$.
ii) Let s_{jg}^R be the set of subjects from imputation cell g with $R_{ij} = 1$ (i.e., Y_{ij} is observed) and s_{jg}^M be the set of subjects from imputation cell g with $R_{ij} = 0$ (i.e., Y_{ij} is missing); let r_{jg} and m_{jg} be the sizes of s_{jg}^R and s_{jg}^M, respectively. Note that $s_{jg} = s_{jg}^R \cup s_{jg}^M$ and $n_{jg} = r_{jg} + m_{jg}$.
iii) If $i \in s_{jg}^M$ (i.e., Y_{ij} is missing), we impute Y_{ij} by $Y_{ij}^\star = Y_{kj}$ where unit k is randomly selected from s_{jg}^R with probability proportional to τ_k. For unweighted random hot-deck imputation, $\tau_k = 1$; for weighted random hot-deck imputation, $\tau_k = w_k = 1/P(k \in s)$.

The subjects in s_{jg}^R are called the *donors* and the subjects in s_{jg}^M are called the *recipients*. Under the proposed procedure, the *donor-recipient* pair is attached to the same cycle. There are three issues related to the above cycle-specific marginal imputation procedure which require some further discussion.

The first issue is the choice of covariates $X_1, \cdots, X_p$ for forming imputation classes. This depends on the variables included in the data set and typically requires some preliminary analysis. All covariates which are significant to the study variable should be included.

The second issue is the presence of time-varying covariates. In this case the imputation cells formed under the proposed procedure may vary from cycle to cycle. This is not a problem in practice but creates some complications for presentations on asymptotic development in the next section. It is always possible, however, to refine and form a single set of imputation cells across all cycles. For instance, suppose there are two cycles, each cycle has two imputation cells, with the first cycle imputation cells formed based on $X_1 = 1$ or $X_1 = 0$ and the second cycle imputation cells formed based on $X_2 = 1$ or $X_2 = 0$. Then a common set of four imputation cells can be used for both cycles, corresponding to $(X_1 = 0, X_2 = 0)$, $(X_1 = 1, X_2 = 0)$, $(X_1 = 0, X_2 = 1)$ and $(X_1 = 1, X_2 = 1)$.

The third issue is that some of the respondent sets s_{jg}^R may be empty when the sample size n is not very large and the number of imputation cells G is not small. This is a common scenario under hot-deck imputation, which is typically handled by collapsing some neighboring cells. More specifically, we can drop some covariates which might not be important or combine adjacent categories of covariates to reduce the total number of cross-classified cells. For asymptotic theory, this is not an issue as the number of cells, G, is assumed fixed as the sample size n gets large.

3. A Pseudo-GEE Method Under Hot-deck Imputation

We assume that the conceptual longitudinal observations $\{(Y_{ij}; X_{ij1}, \cdots, X_{ijp}), j = 1, \cdots, T\}$, $i = 1, \cdots, N$ at the finite population level form a random sample from the superpopulation model ξ as characterized by the following three components:

1. The conditional mean response $\mu_{ij} = E(Y_{ij}|X_{ij})$ is related to the linear predictor $\eta_{ij} = X_{ij}'\boldsymbol{\beta}$ through a monotone link function $g(\cdot)$: $\mu_{ij} = g^{-1}(\eta_{ij}) = g^{-1}(X_{ij}'\boldsymbol{\beta})$, where $X_{ij} = (1, X_{ij1}, \cdots, X_{ijp})'$ and $\boldsymbol{\beta} = (\beta_0, \beta_1, \cdots, \beta_p)'$.
2. The conditional variance of Y_{ij} given X_{ij} is given by $\mathrm{Var}(Y_{ij}|X_{ij}) = \phi v(\mu_{ij})$, where $v(\cdot)$ is the variance function with known form and $\phi > 0$ is called a dispersion parameter.
3. The conditional covariance matrix of $Y_i = (Y_{i1}, \cdots, Y_{iT})'$ is given by $Cov(Y_i|X_i) = A_i^{1/2}\mathbf{R}_i(\alpha)A_i^{1/2}$, where $X_i = (X_{i1}' \cdots, X_{iT}')'$, $A_i = \mathrm{diag}\{\phi v(\mu_{i1}), \cdots, \phi v(\mu_{iT})\}$ and $\mathbf{R}_i(\alpha)$ is the correlation matrix with a specified structure involving parameter α.

   Note that the assumption that the finite population is a random sample from the superpopulation also implies that
4. The response vectors Y_k and Y_l given X_k and X_l are independent for $k \neq l$.

For the estimation procedures described below, we use $\Sigma_i = Cov(Y_i|X_i)$ to denote the true variance-covariance matrix but use V_i to represent the so-called working variance-covariance matrix. In other words, $V_i = A_i^{1/2}\mathbf{R}_i(\alpha)A_i^{1/2}$ when $\mathbf{R}_i(\alpha)$ is a chosen working correlation matrix which does not necessarily coincide with the true one. If the correlation structure is unspecified but is assumed to be constant across all subjects, then $\mathbf{R}_i(\alpha) = \mathbf{R} = (\alpha_{jk})$ can be estimated using the fitted residuals $\boldsymbol{e}_{ij}$; see Equations (3) and (4) in Carrillo et al. (2010) on the estimation of α_{ij} and the dispersion parameter ϕ with complete data. In the presence of missing responses, we only use fitted residuals from observed responses.

Following the generalized estimating equation (GEE) method of Liang and Zeger (1986) and with the complete longitudinal survey sample without any missing values, Carrillo et al. (2010) defined the pseudo-GEE estimator $\hat{\boldsymbol{\beta}}_n$ of the regression coefficients $\boldsymbol{\beta}$ as the solution to $U_n(\boldsymbol{\beta}) = \sum_{i\in s} w_i(\partial \boldsymbol{\mu}_i/\partial \boldsymbol{\beta})' V_i^{-1}(\boldsymbol{y}_i - \boldsymbol{\mu}_i) = \mathbf{0}$, where $\boldsymbol{y}_i$ is the observed value of Y_i and $\boldsymbol{\mu}_i = (\mu_{i1}, \cdots, \mu_{iT})'$.

We now consider the pseudo-GEE method when missing values of the response variable are imputed through the cycle-specific marginal hot-deck imputation method described in Section 2, using either the unweighted or the weighted procedure. Let $\boldsymbol{Y}_i^* = (Y_{i1}^*, \cdots, Y_{iT}^*)'$, where $Y_{ij}^* = Y_{ij}$ if Y_{ij} is observed and $Y_{ij}^* = Y_{ij}^\star$ if Y_{ij} is missing. Let $\boldsymbol{y}_i^*$ be the realized values of Y_i^* from the imputed sample dataset. The pseudo-GEE estimator $\hat{\boldsymbol{\beta}}$ of $\boldsymbol{\beta}$ based on the imputed dataset is defined as the solution to

$$U_n^*(\boldsymbol{\beta}) = \sum_{i\in s} w_i \frac{\partial \boldsymbol{\mu}_i'}{\partial \boldsymbol{\beta}} V_i^{-1}\left(\boldsymbol{y}_i^* - \boldsymbol{\mu}_i\right) = \sum_{g=1}^{G} \sum_{i\in s_g} w_i \frac{\partial \boldsymbol{\mu}_i'}{\partial \boldsymbol{\beta}} V_i^{-1}\left(\boldsymbol{y}_i^* - \boldsymbol{\mu}_i\right) = \mathbf{0}. \tag{1}$$

It can be seen from proofs of major results in the Appendix that all key technical arguments in dealing with the imputed values come from within each imputation cell and then sum up over all cells. For notational simplicity and without loss of generality, we will proceed as if there is only one imputation cell and use s_r and s_m to denote the set of donors and the set of nonrespondents, respectively.

We assume that the conditional distribution of Y given the covariates under the model ξ is independent of the probability sampling design π and the response mechanism R. Under such conditions the model ξ, the design π and the response mechanism R are called unconfounded (Brick et al. 2004). We also assume that all covariates involved in the model ξ are used in forming the imputation classes, and consequently the model ξ and the imputation mechanism (I) are also unconfounded. An important observation under the current setting is that $E_\xi\left(Y_{ij}^\star | X_{ij}\right) = E_\xi(Y_{ij}|X_{ij}) = \mu_{ij}$.

To facilitate the development of variance estimation, for each $i \in s$, we re-arrange the order of components in $\boldsymbol{y}_i^*$ such that $\boldsymbol{y}_i^* = \left(\left(\boldsymbol{y}_i^O\right)', \left(\boldsymbol{y}_i^I\right)'\right)'$, where $\boldsymbol{y}_i^O$ corresponds to the observed Y_{ij}'s and $\boldsymbol{y}_i^I$ denotes the imputed $Y_{ij}^\star$'s. We also have $\boldsymbol{\mu}_i = \left(\left(\boldsymbol{\mu}_i^O\right)', \left(\boldsymbol{\mu}_i^M\right)'\right)'$ arranged in the same order, where $\boldsymbol{\mu}_i^M$ denotes the mean values for the missing Y_{ij}'s. The rows and columns in the working variance-covariance matrix V_i are rotated accordingly.

The consistency of the pseudo-GEE estimator $\hat{\boldsymbol{\beta}}$ is established in Theorem 3.1 below. For a detailed description of the asymptotic framework and the joint randomization approach, see Carrillo et al. (2010). We assume that $r/n \to q \in (0, 1]$ as $n \to \infty$, where r is the number of respondents in the sample. Similar to Theorem 3.1 in Carrillo et al. (2010), the following results are presented in terms of a more general $\psi_i\left(Y_i^*, \boldsymbol{\beta}\right)$ than the specific form $(\partial \boldsymbol{\mu}_i'/\partial \boldsymbol{\beta}) V_i^{-1}\left(\boldsymbol{y}_i^* - \boldsymbol{\mu}_i\right)$ used in the definition of $\hat{\boldsymbol{\beta}}$. The conditioning on X_i is dropped from the notation. The joint randomization involves the model (ξ), the sampling design (π), the response mechanism (R) and the imputation mechanism (I). The response mechanism R is also dropped from the notation since all arguments are conditional on the given set of covariates and the responses are assumed to be missing at random (MAR). The results stated in Theorem 3.1 and the variance estimators developed in the next section are not valid for more general scenarios where the MAR assumption is violated.

Theorem 3.1. *Let $\psi_i(Y_i, \boldsymbol{\beta})$ be an estimating function from $\mathbb{R}^T \times \Theta$ to $\mathbb{R}^p$ with $\Theta \subset \mathbb{R}^p$, and*

$$s_n(\boldsymbol{\beta}) = \sum_{i \in s} w_i \psi_i(Y_i, \boldsymbol{\beta}), \quad s_n^*(\boldsymbol{\beta}) = \sum_{i \in s} w_i \psi_i\left(Y_i^*, \boldsymbol{\beta}\right)$$

be estimating functions for $\boldsymbol{\beta}$ based on complete and imputed data, respectively. Denote $h_i(Y_i) = \sup_{\boldsymbol{\beta} \in \Theta} \|\psi_i(Y_i, \boldsymbol{\beta})\|$, where $\|\cdot\|$ is the usual $\mathscr{L}_1$ norm, and $\Delta_N^*(\boldsymbol{\beta}) = E_{\xi\pi I}\left[N^{-1} s_n^*(\boldsymbol{\beta})\right]$. Suppose that

1. *$\sup_i E_\xi |h_i(Y_i)|^2 < \infty$ and $\sup_i E_\xi \|Y_i\| < \infty$;*
2. *For any $c > 0$ and sequence $\{\boldsymbol{y}_i\}$ satisfying $\|\boldsymbol{y}_i\| \le c$, the sequence of functions $\{g_i(\boldsymbol{\beta}) = \psi_i(\boldsymbol{y}_i, \boldsymbol{\beta})\}$ is equicontinuous on any open subset of Θ;*
3. *The design weights w_i satisfy $N^{-1}\sum_{i \in s} w_i Z_i - N^{-1}\sum_{i=1}^N Z_i = O_p(n^{-1/2})$ for any variable Z such that $N^{-1}\sum_{i=1}^N Z_i^2 = O(1)$;*
4. *The function $\Delta_N(\boldsymbol{\beta}) = E_{\xi\pi}[N^{-1} s_n(\boldsymbol{\beta})]$ satisfies $\Delta_N(\boldsymbol{\beta}_0) = 0$ for some $\boldsymbol{\beta}_0$, and for any $\varepsilon > 0$, there exists a $\delta_\varepsilon > 0$ such that $\inf_{|\boldsymbol{\beta} - \boldsymbol{\beta}_0| > \varepsilon} |\Delta_N(\boldsymbol{\beta})| > \delta_\varepsilon$;*
5. *There is a pseudo-GEE estimator $\hat{\boldsymbol{\beta}} = O_p(1)$ that solves $s_n^*(\boldsymbol{\beta}) = 0$;*

then $\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0 \xrightarrow{p} 0$, where "p" denotes in probability with respect to the model ξ, the sampling design π, the response mechanism R and the imputation mechanism I.

Major steps of the proof of the theorem are outlined in the Appendix. Condition 1 is a moment condition on the superpopulation. Condition 2 requires $\max_i \|g_i(\boldsymbol{\beta}_1) - g_i(\boldsymbol{\beta}_2)\| \to 0$ when $\boldsymbol{\beta}_1 - \boldsymbol{\beta}_2 \to 0$, as long as y_i values are bounded by c. A counter example to Condition 2 can be $\psi(y, \beta) = \beta/y - 1$. Condition 3 is on sampling plan and can be satisfied by most commonly used designs. Condition 4 is an identifiability condition under which $\boldsymbol{\beta}_0$ is uniquely defined.

4. Variance Estimation Under Hot-deck Imputation

In this section we develop linearization variance estimators for the pseudo-GEE estimator $\hat{\boldsymbol{\beta}}$ under the proposed imputation procedure for missing responses. All arguments are conditional on the response mechanism R, i.e., the given pattern of the missing values, which amounts to considering sources of errors due to the model ξ, the sampling design π and the random imputation procedure I.

We present two versions of linearization variance estimators using different routes of approximations. From a theoretical point of view we do not have a clear statement on which variance estimator should be preferred. Results from our simulation studies reported in the next section seem to indicate that the second version has more stable performance under all scenarios we considered in the study.

The first variance estimator we develop follows the conventional route of decomposition of total error into three pieces of errors corresponding to imputation, sampling and model. This is similar to the approach discussed in Särndal (1992). Note that $\hat{\boldsymbol{\beta}}$ is the estimator of $\boldsymbol{\beta}$ based on the imputed dataset, $\hat{\boldsymbol{\beta}}_n$ denotes the estimator based on the complete dataset without missing values, $\boldsymbol{\beta}_N$ represents the census estimator. We have $\hat{\boldsymbol{\beta}} - \boldsymbol{\beta} = (\hat{\boldsymbol{\beta}} - \hat{\boldsymbol{\beta}}_n) + (\hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}_N) + (\boldsymbol{\beta}_N - \boldsymbol{\beta})$. We consider the practical situation for most complex longitudinal surveys where the sampling fraction is small or negligible,

i.e., $n/N = o(1)$. We also assume that the usual $\sqrt{n}$-order applies to $\hat{\boldsymbol{\beta}}$, $\hat{\boldsymbol{\beta}}_n$ and $\boldsymbol{\beta}_N$ so that $\hat{\boldsymbol{\beta}} - \hat{\boldsymbol{\beta}}_n = O_p(1/\sqrt{r})$, $\hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}_N = O_p(1/\sqrt{n})$ and $\boldsymbol{\beta}_N - \boldsymbol{\beta} = O_p(1/\sqrt{N}) = o_p(1/\sqrt{n})$. Under such scenarios we can ignore all terms involving $\boldsymbol{\beta}_N - \boldsymbol{\beta}$. This leads to the following decomposition of the total variance:

$$V_{\text{Tot}} = E_{\xi\pi I}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})' = V_{\text{Imp}} + V_{\text{Sam}} + C_{\text{Imp-Sam}} + C'_{\text{Imp-Sam}} + o(r^{-1}), \quad (2)$$

where $V_{\text{Imp}} = E_{\xi\pi}V_I$, $V_I = E_I(\hat{\boldsymbol{\beta}} - \hat{\boldsymbol{\beta}}_n)(\hat{\boldsymbol{\beta}} - \hat{\boldsymbol{\beta}}_n)'$, $V_{\text{Sam}} = E_\xi V_\pi$, $V_\pi = E_\pi(\hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}_N)(\hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}_N)'$, $C_{\text{Imp-Sam}} = E_{\pi I}C_\xi$ and $C_\xi = E_\xi(\hat{\boldsymbol{\beta}} - \hat{\boldsymbol{\beta}}_n)(\hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}_N)'$. The so-defined total variance V_{Tot} is indeed the mean squared error (MSE) of the estimator $\hat{\boldsymbol{\beta}}$. When the bias of the estimator is negligible, which is the case for the pseudo-GEE estimator in many practical situations, the distinction between variance and MSE vanishes. Let

$$H(\boldsymbol{\beta}) = \sum_{i=1}^{N} \frac{\partial \boldsymbol{\mu}_i'}{\partial \boldsymbol{\beta}} V_i^{-1} \frac{\partial \boldsymbol{\mu}_i}{\partial \boldsymbol{\beta}} \text{ and } \hat{H}(\boldsymbol{\beta}) = \sum_{i \in s} w_i \frac{\partial \boldsymbol{\mu}_i'}{\partial \boldsymbol{\beta}} V_i^{-1} \frac{\partial \boldsymbol{\mu}_i}{\partial \boldsymbol{\beta}},$$

$\pi_i = P(i \in s)$, $\Delta_{ii} = \pi_i(1 - \pi_i)$, $\Delta_{ij} = \pi_{ij} - \pi_i\pi_j$ for $i \neq j$, where $\pi_{ij} = P(i,j \in s)$. Let $A^{\otimes 2}$ denote AA'. Note that $\tau_i = 1$ for unweighted hot-deck imputation and $\tau_i = w_i$ for weighted imputation. Let $s^2_{\tau r} = \sum_{i \in s_r} \tilde{\tau}_i y^2_{ij} - (\bar{y}_{\tau r})^2$, $\bar{y}_{\tau r} = \sum_{i \in s_r} \tilde{\tau}_i \boldsymbol{y}_{ij}$, $\tilde{\tau}_i = \tau_i / \sum_{k \in s_r} \tau_k$, where s_r is the set of donors. Also note that $s^2_{\tau r}$ depends on the specific cycle j under the proposed cycle-specific imputation. The following results are the basis for our first linearization variance estimator.

Theorem 4.1. *Assume that the model ξ, the sampling design π, the response mechanism R and the imputation mechanism I are unconfounded, then the three variance-covariance components can be approximated to order n^{-1} as follows:*

(i) The imputation variance component

$$V_I \approx \left\{\hat{H}(\hat{\boldsymbol{\beta}}_n)\right\}^{-1} \left\{ \sum_{i \in s} w_i^2 \frac{\partial \boldsymbol{\mu}_i'}{\partial \hat{\boldsymbol{\beta}}_n} V_i^{-1} D_i V_i^{-1} \frac{\partial \boldsymbol{\mu}_i}{\partial \hat{\boldsymbol{\beta}}_n} + \left(\sum_{i \in s} w_i \tilde{\boldsymbol{z}}_i \right)^{\otimes 2} \right\} \left\{\hat{H}(\hat{\boldsymbol{\beta}}_n)\right\}^{-1}, \quad (3)$$

where

$$D_i = \begin{pmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \text{diag}\left(s^2_{\tau r}\right) \end{pmatrix}$$

is a $T \times T$ matrix, $\text{diag}\left(s^2_{\tau r}\right)$ *is the diagonal matrix of dimension given by the number of cycles subject i is missing, with diagonal entries given by $s^2_{\tau r}$, $\tilde{\boldsymbol{z}}_i = (\partial \boldsymbol{\mu}_i'/\partial \hat{\boldsymbol{\beta}}_n) V_i^{-1} \left(\left(\boldsymbol{y}_i^O - \boldsymbol{\mu}_i^O \right)', \left(\bar{y}_{\tau r} - \boldsymbol{\mu}_i^M \right)' \right)'$ and $\partial \boldsymbol{\mu}_i / \partial \hat{\boldsymbol{\beta}}_n$ denotes $\partial \boldsymbol{\mu}_i / \partial \boldsymbol{\beta}$ at $\boldsymbol{\beta} = \hat{\boldsymbol{\beta}}_n$.*

(ii) The sampling variance component

$$V_\pi \approx \{H(\boldsymbol{\beta}_N)\}^{-1} \{V_{HTz}\} \{H(\boldsymbol{\beta}_N)\}^{-1}, \quad (4)$$

where the term in the middle, $V_{HTz} = \sum_{i=1}^{N} \sum_{j=1}^{N} \Delta_{ij} (\pi_i \pi_j)^{-1} \boldsymbol{z}_i \boldsymbol{z}_j'$, is the sampling variance of the Horvitz-Thompson estimator of the total of $\boldsymbol{z}_i = (\partial \boldsymbol{\mu}_i'/\partial \boldsymbol{\beta}_N) V_i^{-1} (\boldsymbol{y}_i - \boldsymbol{\mu}_i)$.

(iii) The imputation-sampling covariance component

$$C_\xi \approx \{\hat{H}(\boldsymbol{\beta})\}^{-1}\left\{\sum_{i\in s}\sum_{j\neq i} w_i w_j \frac{\partial \boldsymbol{\mu}_i'}{\partial \boldsymbol{\beta}} V_i^{-1} G_i V_j^{-1} \frac{\partial \boldsymbol{\mu}_j}{\partial \boldsymbol{\beta}} - \sum_{i\in s} w_i^2 \frac{\partial \boldsymbol{\mu}_i'}{\partial \boldsymbol{\beta}} V_i^{-1} K_i V_i^{-1} \frac{\partial \boldsymbol{\mu}_i}{\partial \boldsymbol{\beta}}\right\}\{\hat{H}(\boldsymbol{\beta})\}^{-1}, \tag{5}$$

$$\text{where } G_i = \begin{pmatrix} \mathbf{0} & \mathbf{0} \\ E_\xi\left(\boldsymbol{e}_i^{\mathrm{I}}\left(\boldsymbol{e}_j^{O}\right)'\right) & \mathbf{0} \end{pmatrix} \text{ and } K_i = \begin{pmatrix} \mathbf{0} & \mathbf{0} \\ E_\xi\left(\boldsymbol{e}_i^{M}\left(\boldsymbol{e}_i^{O}\right)'\right) & E_\xi\left(\boldsymbol{e}_i^{M}\left(\boldsymbol{e}_i^{M}\right)'\right) \end{pmatrix}$$

both are $T \times T$ matrix, $\boldsymbol{e}_i^O = \boldsymbol{y}_i^O - \boldsymbol{\mu}_i^O$ and $\boldsymbol{e}_i^I = \boldsymbol{y}_i^I - \boldsymbol{\mu}_i^M$ are the "observed" and "imputed" parts of the error $\boldsymbol{e}_i^* = \boldsymbol{y}_i^* - \boldsymbol{\mu}_i$, respectively, and $\boldsymbol{e}_i^M = \boldsymbol{y}_i^M - \boldsymbol{\mu}_i^M$ is the "missing" part of the error $\boldsymbol{e}_i = \boldsymbol{y}_i - \boldsymbol{\mu}_i$.

Results from Theorem 4.1 enable us to construct a linearization variance estimator through the estimation of the three variance-covariance components (3), (4) and (5). To estimate those three components, the correlation parameters α_{jk} are required for $V_i = A_i^{1/2}\mathbf{R}_i(\alpha)A_i^{1/2}$ and can be estimated using the fitted residuals (Carrillo et al. 2010) from observed responses. The three estimated variance-covariance components are given as follows:

(i) For V_I, we simply replace $\hat{\boldsymbol{\beta}}_n$ by $\hat{\boldsymbol{\beta}}$ and $\boldsymbol{\mu}_i$ by $\hat{\boldsymbol{\mu}}_i$ throughout V_I to obtain $\hat{V}_I$.
(ii) For V_π, we replace $H(\boldsymbol{\beta})$ by $\hat{H}(\hat{\boldsymbol{\beta}})$ and estimate V_{HTz} by

$$\hat{V}_{HTz} = \sum_{i\in s}\sum_{j\in s} \Delta_{ij}(\pi_i\pi_j\pi_{ij})^{-1}\hat{\boldsymbol{z}}_i^*\left(\hat{\boldsymbol{z}}_j^*\right)',$$

where $\hat{\boldsymbol{z}}_i^* = (\partial\boldsymbol{\mu}_i'/\partial\hat{\boldsymbol{\beta}})V_i^{-1}\left(\boldsymbol{y}_i^* - \hat{\boldsymbol{\mu}}_i\right)$. It should be noted that $\boldsymbol{z}_i^* = (\partial\boldsymbol{\mu}_i'/\partial\boldsymbol{\beta}_N)V_i^{-1}\left(\boldsymbol{y}_i^* - \boldsymbol{\mu}_i\right)$ is used for theoretical expressions of the variance V_{HTz} but it cannot be used for the proposed variance estimator because it involves unknown $\boldsymbol{\beta}_N$ and $\boldsymbol{\mu}_i$.

This estimator involves the second order inclusion probabilities π_{ij}. When the finite population sampling fraction is small, it is common practice for survey practitioners to use

$$\hat{V}_{HTz}^* = (n-1)^{-1}\left[n\sum_{k\in s} w_k^2\hat{\boldsymbol{z}}_k^*\left(\hat{\boldsymbol{z}}_k^*\right)'\left(\sum_{i\in s} w_i\hat{\boldsymbol{z}}_i^*\right)^{\otimes 2}\right],$$

which is the variance estimator under the assumption of sampling with-replacement. Unfortunately, due to the use of the imputed dataset instead of a fully observed sample, this quantity has a non-negligible bias as estimator of V_{HTz}. The theoretical expression for

the bias, derived in the appendix under the assumed model ξ, is given by

$$E_\xi\left(\hat{V}^*_{HTz} - V_{HTz}\right) \approx \sum_{k\in s} w_k^2 \frac{\partial \boldsymbol{\mu}_k'}{\partial \boldsymbol{\beta}} V_k^{-1}$$

$$\left[\begin{pmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & E_\xi\left(\boldsymbol{e}_k^{\mathrm{I}}\left(\boldsymbol{e}_k^{\mathrm{I}}\right)'\right) \end{pmatrix} - \begin{pmatrix} \mathbf{0} & E_\xi\left(\boldsymbol{e}_k^{O}\left(\boldsymbol{e}_k^{\mathrm{M}}\right)'\right) \\ E_\xi\left(\boldsymbol{e}_k^{\mathrm{M}}\left(\boldsymbol{e}_k^{O}\right)'\right) & E_\xi\left(\boldsymbol{e}_k^{\mathrm{M}}\left(\boldsymbol{e}_k^{\mathrm{M}}\right)'\right) \end{pmatrix}\right] V_k^{-1} \frac{\partial \boldsymbol{\mu}_k}{\partial \boldsymbol{\beta}} \tag{6}$$

$$- \frac{1}{n-1} \sum_{k\in s}\sum_{i\neq k} w_k w_i \frac{\partial \boldsymbol{\mu}_k'}{\partial \boldsymbol{\beta}} V_k^{-1} \begin{pmatrix} \mathbf{0} & E_\xi\left(\boldsymbol{e}_k^{O}\left(\boldsymbol{e}_k^{\mathrm{I}}\right)'\right) \\ E_\xi\left(\boldsymbol{e}_k^{\mathrm{I}}\left(\boldsymbol{e}_i^{O}\right)'\right) & E_\xi\left(\boldsymbol{e}_k^{\mathrm{I}}\left(\boldsymbol{e}_i^{\mathrm{I}}\right)'\right) \end{pmatrix} V_i^{-1} \frac{\partial \boldsymbol{\mu}_i}{\partial \boldsymbol{\beta}},$$

where $\boldsymbol{e}_k^{\mathrm{I}} = \boldsymbol{y}_k^{\mathrm{I}} - \boldsymbol{\mu}_k^{\mathrm{M}}$, $\boldsymbol{e}_k^{O} = \boldsymbol{y}_k^{O} - \boldsymbol{\mu}_k^{O}$, $\boldsymbol{e}_k^{\mathrm{M}} = \boldsymbol{y}_k^{\mathrm{M}} - \boldsymbol{\mu}_k^{\mathrm{M}}$. We need to estimate this bias to obtain a bias-corrected estimator of V_π.

It is apparent that we only need to obtain estimates for $E_\xi\left(\boldsymbol{e}_k^{I}\left(\boldsymbol{e}_k^{I}\right)'\right)$, $E_\xi\left(\boldsymbol{e}_k^{O}\left(\boldsymbol{e}_k^{I}\right)'\right)$, $E_\xi\left(\boldsymbol{e}_k^{O}\left(\boldsymbol{e}_k^{M}\right)'\right)$ and $E_\xi\left(\boldsymbol{e}_k^{M}\left(\boldsymbol{e}_k^{M}\right)'\right)$, and replace $\boldsymbol{\beta}$ by $\hat{\boldsymbol{\beta}}$. The first two terms can be estimated by $\boldsymbol{r}_k^{I}\left(\boldsymbol{r}_k^{I}\right)'$ and $\boldsymbol{r}_k^{O}\left(\boldsymbol{r}_k^{I}\right)'$, respectively, where $\boldsymbol{r}_k^{O} = \boldsymbol{y}_k^{O} - \hat{\boldsymbol{\mu}}_k^{O}$ and $\boldsymbol{r}_k^{I} = \boldsymbol{y}_k^{I} - \hat{\boldsymbol{\mu}}_k^{\mathrm{M}}$. The last two terms, which involve $\boldsymbol{e}_k^{\mathrm{M}} = \boldsymbol{y}_k^{\mathrm{M}} - \boldsymbol{\mu}_k^{\mathrm{M}}$ with the missing $\boldsymbol{y}_k^{\mathrm{M}}$, cannot be estimated directly.

Our proposed strategy is to first estimate $V = E_\xi(\boldsymbol{e}_i\boldsymbol{e}_i{}') = E_\xi(\boldsymbol{y}_i - \boldsymbol{\mu}_i)(\boldsymbol{y}_i - \boldsymbol{\mu}_i)'$ by $\hat{V}$ using complete cases, and then for each k we partition $\hat{V}$ into $\hat{V}_k^{(OO)}$, $\hat{V}_k^{(OM)}$ and $\hat{V}_k^{(MM)}$ based on the dimensions of $\boldsymbol{y}_k^{O}$ and $\boldsymbol{y}_k^{M}$. We then estimate $E_\xi\left(\boldsymbol{e}_k^{O}\left(\boldsymbol{e}_k^{M}\right)'\right)$ and $E_\xi\left(\boldsymbol{e}_k^{M}\left(\boldsymbol{e}_k^{M}\right)'\right)$ by $\hat{V}_k^{(OM)}$ and $\hat{V}_k^{(MM)}$, respectively. This strategy is similar to estimating the correlation parameters α_{jk} using fitted residuals from observed data. The proposed estimators $\hat{V}_k^{(OM)}$ and $\hat{V}_k^{(MM)}$ should perform reasonably well under the missing-at-random assumption.

(iii) For the estimation of the covariance term C_ξ, the key is to obtain estimates for $E_\xi\left(\boldsymbol{e}_k^{I}\left(\boldsymbol{e}_k^{O}\right)'\right)$, $E_\xi\left(\boldsymbol{e}_k^{O}\left(\boldsymbol{e}_k^{M}\right)'\right)$ and $E_\xi\left(\boldsymbol{e}_k^{M}\left(\boldsymbol{e}_k^{M}\right)'\right)$, which can be handled in the same way described in (ii).

Our first proposed linearization variance estimator is therefore given by

$$\hat{V}_{\mathrm{Tot}} = \hat{V}_I + \hat{V}^*_\pi + \hat{C}_\xi + \hat{C}'_\xi - \widehat{\mathrm{Bias}}\left(\hat{V}^*_\pi\right), \tag{7}$$

where $\hat{V}^*_\pi$ is the estimator of V_π using $\hat{V}^*_{HTz}$, and $\hat{z}^*_i$ as if it was observed; and $\widehat{\mathrm{Bias}}\left(\hat{V}^*_\pi\right)$ is the estimated bias correction term based on (6).

The second variance estimator we develop is based on the assumption that the bias of the pseudo-GEE estimator $\hat{\boldsymbol{\beta}}$ is negligible. Noting that $V_{\mathrm{Tot}} = E_{\xi\pi I}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})' = E_\xi\{E_{\pi I}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})'\} \approx E_\xi\{\mathrm{Var}_{\pi I}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})\}$, we only need to develop an estimator for $\mathrm{Var}_{\pi I}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})$. Since $U_n^*(\hat{\boldsymbol{\beta}}) = \mathbf{0}$ and $\hat{\boldsymbol{\beta}} \xrightarrow{p} \boldsymbol{\beta}$, by using a Taylor series expansion we have

$$\hat{\boldsymbol{\beta}} - \boldsymbol{\beta} = -\left[E_{\pi I}\left(\frac{\partial U_n^*(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}}\right)\right]^{-1} U_n^*(\boldsymbol{\beta}) + o_p(1/\sqrt{r}) = [H(\boldsymbol{\beta})]^{-1} U_n^*(\boldsymbol{\beta}) + o_p(1/\sqrt{r}).$$

It follows that $\mathrm{Var}_{\pi I}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) = [H(\boldsymbol{\beta})]^{-1}\mathrm{Var}_{\pi I}\left[U_n^*(\boldsymbol{\beta})\right][H(\boldsymbol{\beta})]^{-1} + o(1/r)$ and $\mathrm{Var}_{\pi I}\left[U_n^*(\boldsymbol{\beta})\right] = \mathrm{Var}_\pi\left\{E_I\left[U_n^*(\boldsymbol{\beta})\right]\right\} + E_\pi\left\{\mathrm{Var}_I\left[U_n^*(\boldsymbol{\beta})\right]\right\}$. To estimate $\mathrm{Var}_\pi\left\{E_I\left[U_n^*(\boldsymbol{\beta})\right]\right\}$,

we note that

$$E_I\left[U_n^*(\boldsymbol{\beta})\right] = E_I\left[\sum_{i\in s} w_i \frac{\partial \boldsymbol{\mu}_i'}{\partial \boldsymbol{\beta}} V_i^{-1} \begin{pmatrix} \boldsymbol{y}_i^O - \boldsymbol{\mu}_i^O \\ \boldsymbol{y}_i^{\mathrm{I}} - \boldsymbol{\mu}_i^{\mathrm{M}} \end{pmatrix}\right] = \sum_{i\in s} w_i \frac{\partial \boldsymbol{\mu}_i'}{\partial \boldsymbol{\beta}} V_i^{-1} \begin{pmatrix} \boldsymbol{y}_i^O - \boldsymbol{\mu}_i^O \\ \bar{\boldsymbol{y}}_{\pi r} - \boldsymbol{\mu}_i^{\mathrm{M}} \end{pmatrix}$$

which can be written as $E_I[U_n^*(\boldsymbol{\beta})] = \sum_{i\in s} w_i \boldsymbol{z}_{\pi i}$ where $\boldsymbol{z}_{\pi i} = (\partial \boldsymbol{\mu}_i'/\partial \boldsymbol{\beta}) V_i^{-1}\left(\left(\boldsymbol{y}_i^O - \boldsymbol{\mu}_i^O\right)', \left(\bar{\boldsymbol{y}}_{\pi r} - \boldsymbol{\mu}_i^{\mathrm{M}}\right)'\right)'$. Note that $\bar{\boldsymbol{y}}_{\pi r} = \sum_{i\in s_r} \tilde{\tau}_i \boldsymbol{y}_{ij}$ and $\tilde{\tau}_i = \tau_i / \sum_{k\in s_r} \tau_k$.

We could estimate $Var_\pi\left(\sum_{i\in s} w_i \boldsymbol{z}_{\pi i}\right)$ by, say $\hat{V}_{\pi z}$, following the standard variance estimator of a Horvitz-Thompson estimator if $\boldsymbol{z}_{\pi i}$ were observed values and used directly in $\hat{V}_{\pi z}$. The components in $\boldsymbol{z}_{\pi i}$ corresponding to the missing Y_{ij}'s, however, are a constant, i.e., the mean $\bar{\boldsymbol{y}}_{\pi r}$ of the donor set for that particular imputation cell, and there does not seem to be an exact expression for $Var_\pi\left(\sum_{i\in s} w_i \boldsymbol{z}_{\pi i}\right)$ that we could use to derive an approximately unbiased estimator.

Under simple random sampling with a scalar $\boldsymbol{z}_{\pi i}$, the problem reduces to variance estimation for $\bar{z}^* = n^{-1}\left(\sum_{i\in s_r} z_i + \sum_{j\in s_m} z_j^*\right)$ under the mean imputation method, i.e., $z_j^* = \bar{z}_r = r^{-1}\sum_{i\in s_r} z_i$, where r is the number of respondents in the donor set, s_r. For this simple case it can be shown that adjusting the naïve variance estimator $V_{\pi z}$ by the factor $n(n-1)/(r(r-1)) \approx (n/r)^2$ is sufficient, i.e., $(n/r)^2 \hat{V}_{\pi z}$ is an approximately unbiased estimator for $Var_\pi\left(\sum_{i\in s} w_i \boldsymbol{z}_{\pi i}\right)$. The performance of this adjusted variance estimator under general situations with complex survey designs is unknown and requires further exploration.

To estimate $\mathrm{Var}_I\left[U_n^*(\boldsymbol{\beta})\right]$, which is given by

$$\mathrm{Var}_I\left[\sum_{i\in s} w_i \frac{\partial \boldsymbol{\mu}_i'}{\partial \boldsymbol{\beta}} V_i^{-1} \begin{pmatrix} \boldsymbol{y}_i^O - \boldsymbol{\mu}_i^O \\ \boldsymbol{y}_i^{\mathrm{I}} - \boldsymbol{\mu}_i^{\mathrm{M}} \end{pmatrix}\right] = \sum_{i\in s} w_i^2 \frac{\partial \boldsymbol{\mu}_i'}{\partial \boldsymbol{\beta}} V_i^{-1} \begin{pmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathrm{diag}\left(s_{\pi r}^2\right) \end{pmatrix} V_i^{-1} \frac{\partial \boldsymbol{\mu}_i}{\partial \boldsymbol{\beta}}, \tag{8}$$

we only need to replace the model parameter $\boldsymbol{\beta}$ and perhaps also the association parameters by sample based estimates. We denote the resulting estimator of $\mathrm{Var}_I\left[U_n^*(\boldsymbol{\beta})\right]$ as $\hat{V}_{IU}$. Our second proposed alternative variance estimator is given by

$$\hat{V}_A = [\hat{H}(\hat{\boldsymbol{\beta}})]^{-1}\{(n/r)^2 \hat{V}_{\pi z} + \hat{V}_{IU}\}[\hat{H}(\hat{\boldsymbol{\beta}})]^{-1}. \tag{9}$$

5. Simulation Studies

In this section we present results from an extensive simulation study. We used the same superpopulation models and generated the same finite populations as those used in Carrillo et al. (2010), based on a synthetic data file from the first four cycles of NLSCY which was briefly described in Section 1. For continuous response, the finite populations were generated from

$$Y_{ij} = \beta_0 + \beta_1 x_{ij1} + \beta_2 x_{ij1}^2 + \beta_3 x_{ij2} + \beta_4 x_{i3} + \varepsilon_{ij}, \tag{10}$$

where Y_{ij} is the PAS (physical aggression score), x_{ij1} is the AGE, and x_{ij2} is the DeprePMK (depression score of the person most knowledgeable about the child) of subject i at jth cycle; x_{i3} is the GENDER of subject i, $\boldsymbol{\varepsilon}_i = (\varepsilon_{i1}, \varepsilon_{i2}, \varepsilon_{i3}, \varepsilon_{i4}) \sim (\mathbf{0}, \sigma^2 \mathbf{R})$,

$i = 1, \cdots, N = 18,320$, $j = 1, 2, 3, 4$, and $\mathbf{R}$ is the 4×4 correlation matrix. The four covariates in Models (10) and (11) were chosen because they were found to be (the most) significant among the ones analyzed in the previous studies mentioned in Section 1.

For binary response, the finite populations were generated on the basis of the following logistic regression model

$$\text{logit}(p_{ij}) = \beta_0 + \beta_1 x_{ij1} + \beta_2 x_{ij1}^2 + \beta_3 x_{ij2} + \beta_4 x_{i3}, \tag{11}$$

where $p_{ij} = P(Y_{ij} = 1|\mathbf{x}_{ij})$ and $Y_{ij} = 1$ if PAS is high and $Y_{ij} = 0$ if PAS is low. Associations among multi-variate binary response variables were measured by odds ratios.

In order to capture the model (ξ) variability, we regenerated the population response variables Y_{ij} before each simulation sample was selected. For more detailed description of how the model parameters were set and how finite populations were generated from these two models, see Carrillo et al. (2010) and Carrillo-García (2008).

Imputation cells were formed at the finite population level using the three covariates included in the model. To reduce the total number of cells, we collapsed the ordinal variable DeprePMK into three categories. This, together with four categories of age at each cycle and two categories of gender, resulted in a total of 24 imputation cells. Missing responses were randomly "created" on the basis of an overall missing probability p_m ranging from $p_m = 0.05$ to $p_m = 0.25$. Within each imputation cell and for each cycle and for a given overall p_m, the missing probability was set to be q_m which is randomly selected from $\{p_m - 0.02, p_m - 0.01, p_m, p_m + 0.01, p_m + 0.02\}$. In other words, we allowed the missing probabilities to vary a bit from cell to cell and from cycle to cycle, which is most likely the case in practice.

We considered three sampling schemes: (i) simple random sampling (SRS) without replacement; (ii) stratified simple random sampling (STSI); and (iii) cluster sampling with clusters selected by simple random sampling (SIC). Details of the formulation of population strata under STSI and the creation of clusters under SIC are given in Carrillo et al. (2010). The overall sample size used for the simulation under a particular sampling scheme ranges from $n = 120$ to $n = 1{,}200$, and sampling fractions n/N are in between 0.65% and 6.5%. For cluster sampling, the sample sizes are random; the above numbers are expected sample sizes. We obtain, on average, samples ranging from about 5 to 50 elements in each imputation cell. Missing responses were filled using the cycle-specific marginal hot-deck imputation method described in Section 2. Our simulations were programmed in the R software package, as documented in R Development Core Team (2008), and run on a UNIX machine with 24 CPUs. All simulation results were based on 1,000 repeated simulation runs.

We first evaluated the finite sample behavior of the pseudo-GEE estimator $\hat{\boldsymbol{\beta}}$ under the unweighted hot-deck imputation procedure as measured by the simulated relative bias $RB(\hat{\boldsymbol{\beta}}) = 1,000^{-1} \sum_{k=1}^{1,000} (\hat{\boldsymbol{\beta}}^{(k)} - \boldsymbol{\beta})/\boldsymbol{\beta}$, where $\hat{\boldsymbol{\beta}}^{(k)}$ is the estimate of $\boldsymbol{\beta}$ from the kth simulated sample. The simulated results of RB under stratified sampling are presented in Table 2 for the continuous response and in Table 3 for the binary response. Results under other sampling designs can be found in Carrillo-García (2008). Note that $\bar{n}_c$ denotes the average cell sample size and $\bar{r}_c$ is the average number of respondents per imputation cell.

Table 2. Simulated relative bias of $\hat{\boldsymbol{\beta}}$ (in %) for continuous response under STSI

n	$\bar{n}_c$	p_m	$\bar{r}_c$	$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$	$\hat{\beta}_4$
240	10	0.25	7.5	−0.35	−0.57	−0.56	1.36	1.08
		0.20	8.0	0.27	0.40	0.55	0.26	1.15
		0.15	8.5	0.40	0.54	0.63	0.57	1.53
		0.10	9.0	0.35	0.40	0.45	0.41	2.26
		0.05	9.5	0.29	0.28	0.29	0.19	2.32
480	20	0.25	15.0	−0.32	−0.78	−0.97	0.22	2.51
		0.20	16.0	0.16	0.21	0.34	−0.80	1.68
		0.15	17.0	−0.01	−0.02	0.13	−0.55	1.66
		0.10	18.0	0.02	−0.06	−0.01	−0.96	1.69
		0.05	19.0	0.05	0.03	0.11	−0.77	1.31
720	30	0.25	22.5	0.29	0.30	0.29	0.66	2.51
		0.20	24.0	0.02	−0.12	−0.17	0.21	0.10
		0.15	25.5	0.02	−0.13	−0.18	0.51	0.58
		0.10	27.0	0.00	−0.18	−0.23	0.15	0.83
		0.05	28.5	−0.10	−0.33	−0.41	0.39	1.34
960	40	0.25	30.0	0.01	0.03	0.04	−0.29	−0.48
		0.20	32.0	0.03	0.04	0.02	0.32	0.80
		0.15	34.0	0.03	0.07	0.05	0.42	0.23
		0.10	36.0	0.02	−0.01	−0.07	0.18	0.68
		0.05	38.0	−0.01	−0.07	−0.14	−0.05	0.50
1,200	50	0.25	37.5	−0.06	−0.14	−0.16	0.09	−0.40
		0.20	40.0	0.14	0.33	0.47	0.57	0.64
		0.15	42.5	0.21	0.40	0.54	0.65	0.80
		0.10	45.0	0.07	0.18	0.28	0.52	0.60
		0.05	47.5	0.04	0.12	0.21	0.63	0.71

For all three sampling schemes considered and for either continuous or binary response, the biggest relative bias (in absolute value) is about 5%, which occurs with the sample size $n = 240$. In this case there are only around 10 selected subjects per cell. For all other cases the largest relative bias is about 3%, and for sample size $n = 720$ (around 30 per cell) and above, the largest relative bias is bounded by around 2% for all missing fractions considered. While the relative bias tends to decrease as the sample size increases, it does not seem to be influenced by the missing percentages. Estimators of the regression coefficients perform better under stratified sampling for models with continuous responses than for models with binary responses, perhaps due to the fact that stratification is generally less effective for binary responses.

We now turn to the evaluation of performances of variance estimators. We first approximate the true variance-covariance, or more exactly the MSE, matrix of $\hat{\boldsymbol{\beta}}$ under a particular sampling design and a given sample size by $V = 1,000^{-1}\sum_{k=1}^{1,000}(\hat{\boldsymbol{\beta}}^{(k)} - \boldsymbol{\beta})\times(\hat{\boldsymbol{\beta}}^{(k)} - \boldsymbol{\beta})'$ using 1,000 independently simulated samples. The results, not shown here to save space, indicate that many off-diagonal entries of V, corresponding to covariances, are very close to zero. This leads to the following modified definition of relative bias of a variance estimator. Let $\hat{V}$ be an estimator of V; let V_{lm} and $\hat{V}_{lm}$ be the (lm)th entry of V and $\hat{V}$, respectively. The relative bias of $\hat{V}_{lm}$ in estimating V_{lm} based on 1,000 simulated

Table 3. Simulated relative bias of $\hat{\beta}$ (in %) for binary response under STSI

n	$\bar{n}_c$	p_m	$\bar{r}_c$	$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$	$\hat{\beta}_4$
240	10	0.25	7.5	2.01	1.51	1.53	0.12	3.57
		0.20	8.0	1.71	1.43	1.59	0.48	2.96
		0.15	8.5	1.41	1.21	1.34	0.71	2.85
		0.10	9.0	1.65	1.28	1.37	0.17	3.48
		0.05	9.5	1.54	1.20	1.26	0.29	3.61
480	20	0.25	15.0	0.43	0.19	0.10	−1.12	−0.23
		0.20	16.0	−0.12	−0.40	−0.72	0.54	2.65
		0.15	17.0	−0.37	−0.65	−1.04	0.58	2.32
		0.10	18.0	−0.16	−0.45	−0.77	0.45	2.35
		0.05	19.0	−0.19	−0.40	−0.67	0.49	1.43
720	30	0.25	22.5	−0.23	−0.04	−0.06	−0.73	−2.20
		0.20	24.0	−0.01	0.03	−0.15	−0.31	−1.48
		0.15	25.5	0.09	0.08	−0.10	−0.59	−1.61
		0.10	27.0	0.21	0.14	−0.03	−0.89	−0.83
		0.05	28.5	0.39	0.31	0.17	−0.94	−1.23
960	40	0.25	30.0	0.34	0.48	0.55	0.43	−1.40
		0.20	32.0	−0.29	−0.21	−0.33	0.79	−1.20
		0.15	34.0	−0.43	−0.28	−0.38	0.72	−1.34
		0.10	36.0	−0.57	−0.40	−0.54	0.88	−1.47
		0.05	38.0	−0.32	−0.19	−0.25	0.62	−0.99
1,200	50	0.25	37.5	0.28	0.25	0.31	0.38	2.11
		0.20	40.0	0.05	0.11	0.12	0.02	1.53
		0.15	42.5	−0.03	−0.02	−0.01	−0.38	1.32
		0.10	45.0	−0.07	−0.06	−0.07	−0.15	1.46
		0.05	47.5	−0.13	−0.10	−0.11	−0.21	1.27

samples is defined as

$$RB(\hat{V}_{lm}) = \frac{1}{1,000}\sum_{k=1}^{1,000}\left(\hat{V}_{lm}^{(k)} - V_{lm}\right)/(V_{ll}V_{mm})^{1/2},$$

where $\hat{V}_{lm}^{(k)}$ is obtained from the kth simulated sample.

For the linearization variance estimator $\hat{V}_{\text{Tot}}$ given by (7), we included in the simulation several different versions of the estimator to test the importance of each of the variance components and the bias correction term. In one version, we dropped the covariance component $\hat{C}_\xi + \hat{C}'_\xi$; in another version we did not include the bias correction term $\widehat{\text{Bias}}(\hat{V}_\pi^*)$. The version which emerged as a good alternative to $\hat{V}_{\text{Tot}}$ is $\tilde{V}_{\text{Tot}} = \hat{V}_I + \hat{V}_\pi^* + \hat{C}_\xi + \hat{C}'_\xi - \widetilde{\text{Bias}}(\hat{V}_\pi^*)$, where $\widetilde{\text{Bias}}(\hat{V}_\pi^*)$ is a partial bias correction term using only the single summation term in (6) and hence is easier to compute. Simulation results showed that $\hat{V}_{\text{Tot}}$ and $\tilde{V}_{\text{Tot}}$ have very similar performances in almost all cases considered in the simulation and perform better than other versions of variance estimators.

Table 4 presents the simulated relative bias of $\tilde{V}_{\text{Tot}}$ under stratified sampling design for both continuous and binary responses. The three sample sizes, 240, 720 and 1,200, and the three missing probabilities, 0.05, 0.15 and 0.25, represent three scenarios of being small, medium and large considered in the simulation. The relative biases are all

Table 4. Simulated relative bias (in %) of $\tilde{V}_{\mathrm{Tot}}$ under stratified sampling

		Continuous Response					Binary Response				
n	p_m	β_0	β_1	β_2	β_3	β_4	β_0	β_1	β_2	β_3	β_4
240	0.25	−12					−16				
		11	−13				14	−14			
		−11	13	−13			−13	13	−13		
		9	−5	5	−19		−1	5	−4	−18	
		2	2	−2	−1	−11	11	−9	9	−2	−13
	0.15	−11					−7				
		10	−11				6	−6			
		−10	12	−12			−5	5	−4		
		10	−7	7	−19		1	1	−1	−9	
		0	1	−2	2	−3	−1	2	−2	−4	−3
	0.05	−4					0				
		3	−4				0	−1			
		−2	5	−5			0	1	−1		
		10	−8	8	−17		0	4	−4	−6	
		−2	3	−3	4	−1	−4	5	−5	−7	−2
720	0.25	−8					−11				
		7	−7				11	−11			
		−6	6	−6			−11	11	−10		
		−1	4	−4	−11		4	−4	4	−6	
		1	1	−1	−6	−7	−1	4	−4	−2	−13
	0.15	−1					−10				
		1	−3				8	−7			
		−3	5	−6			−6	6	−5		
		1	−1	1	0		0	0	1	−2	
		0	3	−4	0	−4	5	−5	4	2	−5
	0.05	5					−7				
		−4	2				5	−4			
		2	0	−1			−4	3	−2		
		−4	3	−3	4		0	0	0	1	
		0	3	−3	2	−3	5	−5	5	−1	−3
1,200	0.25	−7					−1				
		4	−4				0	0			
		−4	4	−4			1	0	−1		
		5	−2	2	−7		−2	3	−4	−2	
		7	−4	3	−3	−8	2	−1	1	−1	−1
	0.15	−7					3				
		4	−4				−5	5			
		−3	3	−3			5	−4	2		
		4	−1	1	−11		−3	4	−4	0	
		0	−1	1	0	−1	5	−4	2	0	−3
	0.05	−3					7				
		0	1				−8	8			
		1	−1	0			8	−8	6		
		5	−1	1	−8		−4	4	−4	0	
		3	−3	3	−4	1	3	−3	1	3	−1

within 10% when the missing probability is 5% or 15%. When the missing probability is 25%, the relative biases can be as large as 18% for $n = 240$ and binary response, corresponding to $Var(\hat{\beta}_3)$. This is the case where the actual value of the true variance is very small. Relative bias is not a reliable performance measure under such scenarios.

For the second variance estimator $\hat{V}_A = [\hat{H}(\hat{\boldsymbol{\beta}})]^{-1}\{(n/r)^2\hat{V}_{\tau z} + \hat{V}_{IU}\}[\hat{H}(\hat{\boldsymbol{\beta}})]^{-1}$, we also included different versions in the simulation study, especially the one without the inflation factor $(n/r)^2$. Simulation results showed that none of the other versions perform nearly as well as the full version $\hat{V}_A$. Table 5 presents the simulated relative bias of $\hat{V}_A$ under stratified sampling design for both continuous and binary responses. The relative biases are all within 10% for all three missing probabilities and sample sizes considered, except for one entry involving β_3. For a given sample size, the performance of $\hat{V}_A$ does not seem to be influenced by the rates of missing values. In addition, this alternative variance estimator performs much better than the estimator $\hat{V}_{\text{Tot}}$ or $\tilde{V}_{\text{Tot}}$ based on Theorem 4.1.

6. Concluding Remarks

In this article we have proposed a cycle-specific marginal random hot-deck imputation procedure for handling missing responses in longitudinal surveys. We have shown that the pseudo-GEE estimator based on the imputed data set is consistent under certain regularity conditions and the joint randomization framework involving the model, the sampling design, the missing mechanism and the imputation procedure. Two types of linearization variance estimators were developed for the pseudo-GEE estimator of the regression coefficients. Results from an extensive simulation study showed that the proposed imputation procedure works well and the proposed estimators have good finite sample performances.

There are several issues related to the proposed approach. First, the current theoretical development assumes that all covariates used in the GEE model are either categorical or ordinal, and are part of the set of covariates for defining the imputation cells. An obvious problem is that the total number of imputation cells can be very large. This is a common problem in hot-deck imputation and is typically handled by collapsing adjacent cells. The collapsing of cells can also be achieved by dropping some covariates which are not important in terms of modelling the response variables or using reduced number of categories from some covariates as we did in the simulation study. The loss of efficiency, however, is generally unknown under such an ad-hoc procedure. Second, the proposed approach does not extend immediately to the more commonly encountered scenarios where the set of covariates contains both continuous and discrete variables. A possible approach is to first form imputation cells using categorical and ordinal covariates and then carry out the cycle-specific marginal imputation using the nearest-neighbor procedure, with the distance measure defined by the set of continuous covariates. Some preliminary simulation results showed that the pseudo-GEE estimator performs well under this modified approach. Details of the consistency of the pseudo-GEE estimator as well as issues related to variance estimation are currently under investigation. Third, the proposed cycle-specific

Table 5. Simulated relative bias (in %) of $\hat{V}_A$ under stratified sampling

		Continuous Response					Binary Response				
n	p_m	β_0	β_1	β_2	β_3	β_4	β_0	β_1	β_2	β_3	β_4
240	0.25	6					−3				
		−5	5				2	−1			
		4	−4	4			−2	1	0		
		4	−4	4	1		−3	4	−3	−4	
		1	2	−2	0	−4	9	−9	8	−1	−3
	0.15	−4					−2				
		4	−4				1	0			
		−4	5	−6			0	0	1		
		8	−7	7	−10		0	1	−1	−2	
		−1	1	−2	2	0	−1	2	−2	−5	−1
	0.05	2					3				
		−2	1				−3	2			
		2	0	−1			3	−2	3		
		9	−8	7	−12		−1	4	−4	−1	
		−3	3	−3	5	2	−5	5	−5	−7	2
720	0.25	1					−2				
		−2	3				2	−2			
		2	−3	4			−2	2	−1		
		−3	4	−4	0		2	−3	3	2	
		−1	2	−2	−6	−2	−2	4	−3	−2	−7
	0.15	3					−7				
		−2	1				4	−3			
		0	1	−2			−3	2	−1		
		−1	0	1	5		0	−1	1	1	
		−1	3	−3	1	−2	4	−4	4	2	−3
	0.05	4					−6				
		−3	2				4	−3			
		2	0	−1			−3	2	−2		
		−4	3	−3	5		0	0	0	2	
		0	3	−4	2	−3	5	−5	5	−1	−2
1,200	0.25	2					6				
		−4	5				−7	7			
		5	−5	5			7	−7	7		
		2	−1	1	2		−2	2	−2	5	
		6	−4	3	−3	−4	0	0	0	0	3
	0.15	−3					5				
		1	0				−7	7			
		−1	0	0			7	−6	5		
		3	−1	1	−7		−4	3	−4	4	
		0	−1	1	0	1	4	−3	2	0	−1
	0.05	−2					7				
		−1	2				−8	9			
		1	−1	0			8	−8	6		
		5	−1	1	−7		−3	4	−4	1	
		3	−3	3	−4	1	3	−3	2	3	−1

marginal imputation procedure ignores the correlation structure among the longitudinal measurements. This may not be an issue when the rates of missing data are small, say less than 25%. When the rates are large, more sophisticated imputation procedures which preserve the correlation structure are clearly desirable. Fourth, for the estimation of the association parameter α in the GEE model, it is recommended that only complete cases be used instead of the imputed dataset, due to the nature of the proposed imputation procedure.

There are several directions in which the current work can be extended. First, for simulation studies, it would be of interest to compare the imputation approach proposed here to a probability re-weighting method similar to Robins et al. (1995) but with survey weights incorporated. Second, it is of both theoretical and practical interest to extend the method to cases where both the response variables and the covariates are subject to missingness. Third, it would be of great interest to survey practitioners and longitudinal survey data users to develop replication weights for variance estimation which provides valid results for GEE analysis under the proposed imputation procedure.

Appendix: Proofs

Proof of Theorem 3.1. Proof of consistency of the pseudo-GEE estimator $\hat{\boldsymbol{\beta}}$ requires the following lemma. A proof of the lemma can be found in Carrillo-García (2008).

Lemma 3.1. *Let Y_i^* be the vector Y_i with missing values imputed by the hot-deck method (weighted or unweighted); suppose that Θ is a compact subset of $\mathbb{R}^p$ and that conditions 1, 2, and 3 in Theorem 3.1 hold, then, as $r, n, N \to \infty$,*

$$\sup_{\boldsymbol{\beta}\in\Theta}\left|\left|\frac{1}{N}s_n^*(\boldsymbol{\beta}) - \Delta_N^*(\boldsymbol{\beta})\right|\right| \xrightarrow{p} 0,$$

where $s_n^*(\boldsymbol{\beta}) = \sum_{i\in s} w_i\psi_i\left(Y_i^*, \boldsymbol{\beta}\right)$ and $\Delta_N^*(\boldsymbol{\beta}) = N^{-1}E_{\xi\pi RI}\left[\sum_{i\in s} w_i\psi_i\left(Y_i^*, \boldsymbol{\beta}\right)\right]$.

Without loss of generality we consider cases where $\psi_i\left(Y_i^*, \boldsymbol{\beta}\right)$ is linear in Y_i^*, which implies that, under the proposed imputation procedure, $E_{\xi I}\left[\psi_i\left(Y_i^*, \boldsymbol{\beta}\right)\right] = E_{\xi I}[\psi_i(Y_i, \boldsymbol{\beta})]$. This further implies that, for any $\boldsymbol{\beta} \in \Theta$,

$$\Delta_N^*(\boldsymbol{\beta}) = E_{\pi R}\left[\frac{1}{N}\sum_{i\in s} w_i E_{\xi I}\left[\psi_i\left(Y_i^*, \boldsymbol{\beta}\right)\right]\right] = E_{\pi R}\left[\frac{1}{N}\sum_{i\in s} w_i E_{\xi I}\left[\psi_i\left(Y_i, \boldsymbol{\beta}\right)\right]\right] = \Delta_N(\boldsymbol{\beta}).$$

The rest of the proof follows the same lines of the proof of Theorem 3.1 in Carrillo et al. (2010) for the case of complete responses. □

Proof of Theorem 4.1. We begin with V_I, the variance component due to imputation. Noting that $\hat{\boldsymbol{\beta}}$ solves $U_n^*(\hat{\boldsymbol{\beta}}) = \mathbf{0}$ and $\hat{\boldsymbol{\beta}} = \hat{\boldsymbol{\beta}}_n + o_p(1)$, we can apply Taylor series

expansion to $U_n^*(\hat{\boldsymbol{\beta}})$ around $\hat{\boldsymbol{\beta}}_n$ to obtain

$$\begin{aligned} U_n^*(\hat{\boldsymbol{\beta}}) &= U_n^*(\hat{\boldsymbol{\beta}}_n) + \frac{\partial U_n^*(\hat{\boldsymbol{\beta}}_n)}{\partial \hat{\boldsymbol{\beta}}_n}(\hat{\boldsymbol{\beta}} - \hat{\boldsymbol{\beta}}_n) + o_p(N/\sqrt{r}) \\ &= U_n^*(\hat{\boldsymbol{\beta}}_n) + E_I\left(\frac{\partial U_n^*(\hat{\boldsymbol{\beta}}_n)}{\partial \hat{\boldsymbol{\beta}}_n}\right)(\hat{\boldsymbol{\beta}} - \hat{\boldsymbol{\beta}}_n) \\ &\quad + \left[\frac{\partial U_n^*(\hat{\boldsymbol{\beta}}_n)}{\partial \hat{\boldsymbol{\beta}}_n} - E_I\left(\frac{\partial U_n^*(\hat{\boldsymbol{\beta}}_n)}{\partial \hat{\boldsymbol{\beta}}_n}\right)\right](\hat{\boldsymbol{\beta}} - \hat{\boldsymbol{\beta}}_n) + o_p(N/\sqrt{r}) \\ &= U_n^*(\hat{\boldsymbol{\beta}}_n) + E_I\left(\frac{\partial U_n^*(\hat{\boldsymbol{\beta}}_n)}{\partial \hat{\boldsymbol{\beta}}_n}\right)(\hat{\boldsymbol{\beta}} - \hat{\boldsymbol{\beta}}_n) + o_p(N/\sqrt{r}). \end{aligned}$$

Furthermore, we have

$$\begin{aligned} \frac{\partial U_n^*(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}} &= \frac{\partial}{\partial \boldsymbol{\beta}}\left[\sum_{i\in s} w_i \frac{\partial \boldsymbol{\mu}_i'}{\partial \boldsymbol{\beta}} V_i^{-1}\left(\mathbf{y}_i^* - \boldsymbol{\mu}_i\right)\right] \\ &= \sum_{i\in s} w_i \frac{\partial}{\partial \boldsymbol{\beta}}\left(\frac{\partial \boldsymbol{\mu}_i'}{\partial \boldsymbol{\beta}} V_i^{-1}\right)\left(\mathbf{y}_i^* - \boldsymbol{\mu}_i\right) - \sum_{i\in s} w_i \frac{\partial \boldsymbol{\mu}_i{}'}{\partial \boldsymbol{\beta}} V_i^{-1} \frac{\partial \boldsymbol{\mu}_i}{\partial \boldsymbol{\beta}} \\ &= -\sum_{i\in s} w_i \frac{\partial \boldsymbol{\mu}_i'}{\partial \boldsymbol{\beta}} V_i^{-1} \frac{\partial \boldsymbol{\mu}_i}{\partial \boldsymbol{\beta}} + O_p(N/\sqrt{r}). \end{aligned}$$

This leads to $E_I\left(\partial U_n^*(\hat{\boldsymbol{\beta}}_n)/\partial \hat{\boldsymbol{\beta}}_n\right) = -\hat{H}(\hat{\boldsymbol{\beta}}_n) + O_p(N/\sqrt{r})$. It follows that $\hat{\boldsymbol{\beta}} - \hat{\boldsymbol{\beta}}_n = [\hat{H}(\hat{\boldsymbol{\beta}}_n)]^{-1} U_n^*(\hat{\boldsymbol{\beta}}_n) + o_p(1/\sqrt{r})$, and

$$\begin{aligned} V_I &= E_I(\hat{\boldsymbol{\beta}} - \hat{\boldsymbol{\beta}}_n)(\hat{\boldsymbol{\beta}} - \hat{\boldsymbol{\beta}}_n)' \\ &= [\hat{H}(\hat{\boldsymbol{\beta}}_n)]^{-1} E_I\left[U_n^*(\hat{\boldsymbol{\beta}}_n)\left(U_n^*(\hat{\boldsymbol{\beta}}_n)\right)'\right]\left[\hat{H}(\hat{\boldsymbol{\beta}}_n)\right]^{-1} + o_p\left(\frac{1}{r}\right). \end{aligned} \tag{12}$$

Under the proposed random hot-deck imputation, either unweighted ($\tau_j = 1$) or weighted ($\tau_j = w_j$), and using the simplified notation s_r for the set of donors and s_m for the set of recipients, we have, for $j \in s_m$, $E_I\left(y_j^*\right) = \sum_{i\in s_r} \tau_i y_i / \sum_{i\in s_r} \tau_i = \bar{y}_{\tau r}$ and

$$\mathrm{Var}_I\left(y_j^*\right) = \frac{\sum_{i\in s_r} \tau_i y_i^2}{\sum_{i\in s_r} \tau_i} - \left(\frac{\sum_{i\in s_r} \tau_i y_i}{\sum_{i\in s_r} \tau_i}\right)^2 = s_{\tau r}^2$$

In addition, $\boldsymbol{y}_j^*$ and $\boldsymbol{y}_k^*$ are independent if $j \neq k$. It follows that

$$
\begin{aligned}
E_I\left(\boldsymbol{y}_i^* - \boldsymbol{\mu}_i\right)\left(\boldsymbol{y}_i^* - \boldsymbol{\mu}_i\right)' &= \mathrm{Var}_I\left(\boldsymbol{y}_i^* - \boldsymbol{\mu}_i\right) + E_I\left(\boldsymbol{y}_i^* - \boldsymbol{\mu}_i\right)E_I\left(\boldsymbol{y}_i^* - \boldsymbol{\mu}_i\right)' \\
&= \mathrm{Var}_I\left(\boldsymbol{y}_i^*\right) + \left(E_I\boldsymbol{y}_i^* - \boldsymbol{\mu}_i\right)\left(E_I\boldsymbol{y}_i^* - \boldsymbol{\mu}_i\right)' \\
&= \mathrm{Var}_I\begin{pmatrix}\boldsymbol{y}_i^O \\ \boldsymbol{y}_i^{\mathrm{I}}\end{pmatrix} + \left[E_I\begin{pmatrix}\boldsymbol{y}_i^O \\ \boldsymbol{y}_i^{\mathrm{I}}\end{pmatrix} - \boldsymbol{\mu}_i\right]\left[E_I\begin{pmatrix}\boldsymbol{y}_i^O \\ \boldsymbol{y}_i^{\mathrm{I}}\end{pmatrix} - \boldsymbol{\mu}_i\right]' \\
&= \begin{pmatrix}\mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathrm{diag}(s_{\tau r}^2)\end{pmatrix} + \left[\begin{pmatrix}\boldsymbol{y}_i^O \\ \bar{\boldsymbol{y}}_{\tau r}\end{pmatrix} - \boldsymbol{\mu}_i\right]\left[\begin{pmatrix}\boldsymbol{y}_i^O \\ \bar{\boldsymbol{y}}_{\tau r}\end{pmatrix} - \boldsymbol{\mu}_i\right]'
\end{aligned}
$$

and

$$
\begin{aligned}
E_I\left(\boldsymbol{y}_i^* - \boldsymbol{\mu}_i\right)\left(\boldsymbol{y}_j^* - \boldsymbol{\mu}_j\right)' &= \left[E_I\begin{pmatrix}\boldsymbol{y}_i^O \\ \boldsymbol{y}_i^{\mathrm{I}}\end{pmatrix} - \boldsymbol{\mu}_i\right]\left[E_I\begin{pmatrix}\boldsymbol{y}_j^O \\ \boldsymbol{y}_j^{\mathrm{I}}\end{pmatrix} - \boldsymbol{\mu}_j\right]' \\
&= \left[\begin{pmatrix}\boldsymbol{y}_i^O \\ \bar{\boldsymbol{y}}_{\tau r}\end{pmatrix} - \boldsymbol{\mu}_i\right]\left[\begin{pmatrix}\boldsymbol{y}_j^O \\ \bar{\boldsymbol{y}}_{\tau r}\end{pmatrix} - \boldsymbol{\mu}_j\right]'
\end{aligned}
$$

for $i \neq j$. This further leads to the following expressions for $E_I\left[U_n^*(\hat{\boldsymbol{\beta}}_n)(U_n^*(\hat{\boldsymbol{\beta}}_n))'\right]$:

$$
\begin{aligned}
&E_I\left[\sum_{i\in s} w_i \frac{\partial \boldsymbol{\mu}_i'}{\partial \hat{\boldsymbol{\beta}}_n} V_i^{-1}\left(\boldsymbol{y}_i^* - \boldsymbol{\mu}_i\right) \cdot \sum_{i\in s} w_i\left(\boldsymbol{y}_i^* - \boldsymbol{\mu}_i\right)' V_i^{-1} \frac{\partial \boldsymbol{\mu}_i}{\partial \hat{\boldsymbol{\beta}}_n}\right] \\
&= E_I\left[\sum_{i\in s} w_i^2 \frac{\partial \boldsymbol{\mu}_i'}{\partial \hat{\boldsymbol{\beta}}_n} V_i^{-1}\left(\boldsymbol{y}_i^* - \boldsymbol{\mu}_i\right)\left(\boldsymbol{y}_i^* - \boldsymbol{\mu}_i\right)' V_i^{-1} \frac{\partial \boldsymbol{\mu}_i}{\partial \hat{\boldsymbol{\beta}}_n}\right. \\
&\quad \left. + \sum_{i,j\in s}\sum_{i\neq j} w_i w_j \frac{\partial \boldsymbol{\mu}_i'}{\partial \hat{\boldsymbol{\beta}}_n} V_i^{-1}\left(\boldsymbol{y}_i^* - \boldsymbol{\mu}_i\right)\left(\boldsymbol{y}_j^* - \boldsymbol{\mu}_j\right)' V_j^{-1} \frac{\partial \boldsymbol{\mu}_j}{\partial \hat{\boldsymbol{\beta}}_n}\right] \\
&= \sum_{i\in s} w_i^2 \frac{\partial \boldsymbol{\mu}_i'}{\partial \hat{\boldsymbol{\beta}}_n} V_i^{-1} \begin{pmatrix}\mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathrm{diag}\left(s_{\tau r}^2\right)\end{pmatrix} V_i^{-1} \frac{\partial \boldsymbol{\mu}_i}{\partial \hat{\boldsymbol{\beta}}_n} \\
&\quad + \sum_{i\in s} w_i^2 \frac{\partial \boldsymbol{\mu}_i'}{\partial \hat{\boldsymbol{\beta}}_n} V_i^{-1} \left[\begin{pmatrix}\boldsymbol{y}_i^O \\ \bar{\boldsymbol{y}}_{\tau r}\end{pmatrix} - \boldsymbol{\mu}_i\right]^{\otimes 2} V_i^{-1} \frac{\partial \boldsymbol{\mu}_i}{\partial \hat{\boldsymbol{\beta}}_n}
\end{aligned}
$$

$$
\begin{aligned}
&+\sum_{i,j\in s}\sum_{i\neq j} w_i w_j \frac{\partial \boldsymbol{\mu}_i'}{\partial \hat{\boldsymbol{\beta}}_n} V_i^{-1}\left[\begin{pmatrix} \boldsymbol{y}_i^O \\ \bar{\boldsymbol{y}}_{\pi r}\end{pmatrix} - \boldsymbol{\mu}_i\right]\left[\begin{pmatrix} \boldsymbol{y}_j^O \\ \bar{\boldsymbol{y}}_{\pi r}\end{pmatrix} - \boldsymbol{\mu}_j\right]' V_j^{-1}\frac{\partial \boldsymbol{\mu}_j}{\partial \hat{\boldsymbol{\beta}}_n} \\
&= \sum_{i\in s} w_i^2 \frac{\partial \boldsymbol{\mu}_i'}{\partial \hat{\boldsymbol{\beta}}_n} V_i^{-1}\begin{pmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \operatorname{diag}\left(s_{\pi r}^2\right)\end{pmatrix} V_i^{-1}\frac{\partial \boldsymbol{\mu}_i}{\partial \hat{\boldsymbol{\beta}}_n} \\
&+\sum_{i\in s}\sum_{j\in s} w_i w_j \frac{\partial \boldsymbol{\mu}_i'}{\partial \hat{\boldsymbol{\beta}}_n} V_i^{-1}\left[\begin{pmatrix} \boldsymbol{y}_i^O \\ \bar{\boldsymbol{y}}_{\pi r}\end{pmatrix} - \boldsymbol{\mu}_i\right]\left[\begin{pmatrix} \boldsymbol{y}_j^O \\ \bar{\boldsymbol{y}}_{\pi r}\end{pmatrix} - \boldsymbol{\mu}_j\right]' V_j^{-1}\frac{\partial \boldsymbol{\mu}_j}{\partial \hat{\boldsymbol{\beta}}_n} \qquad (13) \\
&= \sum_{i\in s} w_i^2 \frac{\partial \boldsymbol{\mu}_i'}{\partial \hat{\boldsymbol{\beta}}_n} V_i^{-1}\begin{pmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \operatorname{diag}\left(s_{\pi r}^2\right)\end{pmatrix} V_i^{-1}\frac{\partial \boldsymbol{\mu}_i}{\partial \hat{\boldsymbol{\beta}}_n} \\
&+\left[\sum_{i\in s} w_i \frac{\partial \boldsymbol{\mu}_i'}{\partial \hat{\boldsymbol{\beta}}_n} V_i^{-1}\begin{pmatrix} \boldsymbol{y}_i^O - \boldsymbol{\mu}_i^O \\ \bar{\boldsymbol{y}}_{\pi r} - \boldsymbol{\mu}_i^{\mathrm{M}}\end{pmatrix}\right]^{\otimes 2}.
\end{aligned}
$$

This completes the derivation of V_I. As for V_π, it is a direct consequence of the expansion $\hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}_N = [H(\boldsymbol{\beta}_N)]^{-1}U_n(\boldsymbol{\beta}_N) + o_p(1/\sqrt{n})$. The expression for C_ξ is derived based on the two expansions for $\hat{\boldsymbol{\beta}} - \hat{\boldsymbol{\beta}}_n$ and $\hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}_N$. Details are omitted. □

Derivation of the bias term given by (6). First, we have

$$
\begin{aligned}
z_k^* z_k^{*\prime} - z_k z_k' &= \frac{\partial \boldsymbol{\mu}_k'}{\partial \boldsymbol{\beta}_N} V_k^{-1}\left[\begin{pmatrix} \boldsymbol{e}_k^O \\ \boldsymbol{e}_k^{\mathrm{I}}\end{pmatrix}^{\otimes 2} - \begin{pmatrix} \boldsymbol{e}_k^O \\ \boldsymbol{e}_k^{\mathrm{M}}\end{pmatrix}^{\otimes 2}\right] V_k^{-1}\frac{\partial \boldsymbol{\mu}_k}{\partial \boldsymbol{\beta}_N} \\
&= \frac{\partial \boldsymbol{\mu}_k'}{\partial \boldsymbol{\beta}_N} V_k^{-1}\left[\begin{pmatrix} \mathbf{0} & \boldsymbol{e}_k^O \boldsymbol{e}_k^{\mathrm{I}'} \\ \boldsymbol{e}_k^{\mathrm{I}} \boldsymbol{e}_k^{O'} & \boldsymbol{e}_k^{\mathrm{I}} \boldsymbol{e}_k^{\mathrm{I}'}\end{pmatrix} - \begin{pmatrix} \mathbf{0} & \boldsymbol{e}_k^O \boldsymbol{e}_k^{\mathrm{M}'} \\ \boldsymbol{e}_k^{\mathrm{M}} \boldsymbol{e}_k^{O'} & \boldsymbol{e}_k^{\mathrm{M}} \boldsymbol{e}_k^{\mathrm{M}'}\end{pmatrix}\right] V_k^{-1}\frac{\partial \boldsymbol{\mu}_k}{\partial \boldsymbol{\beta}_N}
\end{aligned}
$$

and

$$
\begin{aligned}
z_k^* z_i^{*\prime} - z_k z_i' &= \frac{\partial \boldsymbol{\mu}_k'}{\partial \boldsymbol{\beta}_N} V_k^{-1}\left[\begin{pmatrix} \boldsymbol{e}_k^O \\ \boldsymbol{e}_k^{\mathrm{I}}\end{pmatrix}\left(\boldsymbol{e}_i^{O'}, \boldsymbol{e}_i^{\mathrm{I}'}\right) - \begin{pmatrix} \boldsymbol{e}_k^O \\ \boldsymbol{e}_k^{\mathrm{M}}\end{pmatrix}\left(\boldsymbol{e}_i^{O'}, \boldsymbol{e}_i^{\mathrm{M}'}\right)\right] V_i^{-1}\frac{\partial \boldsymbol{\mu}_i}{\partial \boldsymbol{\beta}_N} \\
&= \frac{\partial \boldsymbol{\mu}_k'}{\partial \boldsymbol{\beta}_N} V_k^{-1}\left[\begin{pmatrix} \mathbf{0} & \boldsymbol{e}_k^O \boldsymbol{e}_i^{\mathrm{I}'} \\ \boldsymbol{e}_k^{\mathrm{I}} \boldsymbol{e}_i^{O'} & \boldsymbol{e}_k^{\mathrm{I}} \boldsymbol{e}_i^{\mathrm{I}'}\end{pmatrix} - \begin{pmatrix} \mathbf{0} & \boldsymbol{e}_k^O \boldsymbol{e}_i^{\mathrm{M}'} \\ \boldsymbol{e}_k^{\mathrm{M}} \boldsymbol{e}_i^{O'} & \boldsymbol{e}_k^{\mathrm{M}} \boldsymbol{e}_i^{\mathrm{M}'}\end{pmatrix}\right] V_i^{-1}\frac{\partial \boldsymbol{\mu}_i}{\partial \boldsymbol{\beta}_N}
\end{aligned}
$$

for $i \neq k$. This leads to

$$E_\xi\left[\boldsymbol{z}_k^*\boldsymbol{z}_k^{*\prime} - \boldsymbol{z}_k\boldsymbol{z}_k'\right] = \frac{\partial \boldsymbol{\mu}_k'}{\partial \boldsymbol{\beta}_N} V_k^{-1} \left[\begin{pmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & E_\xi\left(\boldsymbol{e}_k^{\mathrm{I}}\boldsymbol{e}_k^{\mathrm{I}\prime}\right) \end{pmatrix} - \begin{pmatrix} \mathbf{0} & E_\xi\left(\boldsymbol{e}_k^{O}\boldsymbol{e}_k^{\mathrm{M}\prime}\right) \\ E_\xi\left(\boldsymbol{e}_k^{\mathrm{M}}\boldsymbol{e}_k^{O\prime}\right) & E_\xi\left(\boldsymbol{e}_k^{\mathrm{M}}\boldsymbol{e}_k^{\mathrm{M}\prime}\right) \end{pmatrix} \right] V_k^{-1} \frac{\partial \boldsymbol{\mu}_k}{\partial \boldsymbol{\beta}_N} \tag{14}$$

and

$$E_\xi\left[\boldsymbol{z}_k^*\boldsymbol{z}_i^{*\prime} - \boldsymbol{z}_k\boldsymbol{z}_i'\right] = \frac{\partial \boldsymbol{\mu}_k'}{\partial \boldsymbol{\beta}_N} V_k^{-1} \begin{pmatrix} \mathbf{0} & E_\xi\left(\boldsymbol{e}_k^{O}\boldsymbol{e}_i^{\mathrm{I}'}\right) \\ E_\xi\left(\boldsymbol{e}_k^{\mathrm{I}}\boldsymbol{e}_i^{O'}\right) & E_\xi\left(\boldsymbol{e}_k^{\mathrm{I}}\boldsymbol{e}_i^{\mathrm{I}\prime}\right) \end{pmatrix} V_i^{-1} \frac{\partial \boldsymbol{\mu}_i}{\partial \boldsymbol{\beta}_N}. \tag{15}$$

for $i \neq k$. The model-based bias $E_\xi\left(\hat{V}_{HTz}^* - V_{HTz}\right)$ (multiplied by the factor $n-1$) is therefore given by

$$E_\xi\left\{ \left[(n-1)\sum_{k\in s} w_k^2 \boldsymbol{z}_k^*\boldsymbol{z}_k^{*\prime} - \sum_{k\in s}\sum_{i\neq k} w_k w_i \boldsymbol{z}_k^*\boldsymbol{z}_i^{*\prime} \right] - \left[(n-1)\sum_{k\in s} w_k^2 \boldsymbol{z}_k\boldsymbol{z}_k' - \sum_{k\in s}\sum_{i\neq k} w_k w_i \boldsymbol{z}_k\boldsymbol{z}_i' \right] \right\}$$

$$= E_\xi\left\{ (n-1)\sum_{k\in s} w_k^2 \left(\boldsymbol{z}_k^*\boldsymbol{z}_k^{*\prime} - \boldsymbol{z}_k\boldsymbol{z}_k'\right) - \sum_{k\in s}\sum_{i\neq k} w_k w_i \left(\boldsymbol{z}_k^*\boldsymbol{z}_i^{*\prime} - \boldsymbol{z}_k\boldsymbol{z}_i'\right) \right\}$$

$$= (n-1)\sum_{k\in s} w_k^2 \frac{\partial \boldsymbol{\mu}_k'}{\partial \boldsymbol{\beta}_N} V_k^{-1} \left[\begin{pmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & E_\xi\left(\boldsymbol{e}_k^{\mathrm{I}}\boldsymbol{e}_k^{\mathrm{I}\prime}\right) \end{pmatrix} - \begin{pmatrix} \mathbf{0} & E_\xi\left(\boldsymbol{e}_k^{O}\boldsymbol{e}_k^{\mathrm{M}\prime}\right) \\ E_\xi\left(\boldsymbol{e}_k^{\mathrm{M}}\boldsymbol{e}_k^{O\prime}\right) & E_\xi\left(\boldsymbol{e}_k^{\mathrm{M}}\boldsymbol{e}_k^{\mathrm{M}\prime}\right) \end{pmatrix} \right] V_k^{-1} \frac{\partial \boldsymbol{\mu}_k}{\partial \boldsymbol{\beta}_N}$$

$$- \sum_{k\in s}\sum_{i\neq k} w_k w_i \frac{\partial \boldsymbol{\mu}_k'}{\partial \boldsymbol{\beta}_N} V_k^{-1} \begin{pmatrix} \mathbf{0} & E_\xi\left(\boldsymbol{e}_k^{O}\boldsymbol{e}_i^{\mathrm{I}\prime}\right) \\ E_\xi\left(\boldsymbol{e}_k^{\mathrm{I}}\boldsymbol{e}_i^{O\prime}\right) & E_\xi\left(\boldsymbol{e}_k^{\mathrm{I}}\boldsymbol{e}_i^{\mathrm{I}\prime}\right) \end{pmatrix} V_i^{-1} \frac{\partial \boldsymbol{\mu}_i}{\partial \boldsymbol{\beta}_N}.$$

□

7. References

Binder, D.A. and Patak, Z. (1994). Use of Estimating Functions for Estimation from Complex Surveys. Journal of the American Statistical Association, 89, 1035–1043.

Brick, J.M., Kalton, G., and Kim, J.K. (2004). Variance Estimation with Hot Deck Imputation using a Model. Survey Methodology, 30, 57–66.

Carrillo, I., Chen, J., and Wu, C. (2010). The Pseudo-GEE Approach to the Analysis of Longitudinal Surveys. The Canadian Journal of Statistics/La Revue Canadienne de Statistique, 38, 540–554.

Carrillo, I., Chu, C., Su, W., and Xie, X. (2005). A Longitudinal Study of Factors Affecting Children's Behaviour. Proceedings of the Survey Methods Section, Saskatoon. Statistical Society of Canada.

Carrillo-García, I.A. (2006). Analysis of Longitudinal Survey Data with Missing Observations: An Application of Weighted Gee to the National Longitudinal Survey of

Children and Youth (NLSCY). Technical report, Statistics Canada. MITACS/NPCDS Intern ship Program.

Carrillo-García, I.A. (2008). Analysis of Longitudinal Surveys with Missing Responses. PhD thesis, University of Waterloo, ON, Canada.

Diggle, P., Heagerty, P., Liang, K.-Y., and Zeger, S. (2002). Analysis of Longitudinal Data (Second Edition). New York: Oxford University Press.

Ford, B.M. (1983). Incomplete Data in Sample Surveys, Volume 2. Chapter An Overview of Hot-deck Procedures, 185–207, New York: Academic Press.

Godambe, V.P. (1995). Estimation of Parameters in Survey Sampling: Optimality. The Canadian Journal of Statistics/La Revue Canadienne de Statistique, 23, 227–243.

Godambe, V.P. and Thompson, M.E. (1986). Parameters of Superpopulation and Survey Population: Their Relationships and Estimation. International Statistical Review, 54, 127–138.

Groves, R.M., Dillman, D.A., Eltinge, J.L. and Little, R.J.A. (Eds) (2002). Survey Nonresponse. Wiley Series in Probability and Statistics. New York: John Wiley & Sons.

Hedeker, D. and Gibbons, R.D. (2006). Longitudinal Data Analysis. Wiley Series in Probability and Statistics. Hoboken: John Wiley & Sons.

Liang, K.-Y. and Zeger, S.L. (1986). Longitudinal Data Analysis using Generalized Linear Models. Biometrika, 73, 13–22.

R Development Core Team (2008). R: A Language and Environment for Statistical Computing. Vienna, Austria: R Foundation for Statistical Computing, ISBN 3-900051-07-0.

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, 106–121.

Rubin-Bleuer, S. and Schiopu Kratina, I. (2005). On the Two-phase Framework for Joint Model and Design-based Inference. The Annals of Statistics, 33, 2789–2810.

Sande, I.G. (1983). Incomplete Data in Sample Surveys, Volume 3. Chapter Hot-deck Imputation Procedures, 339–349, New York: Academic Press.

Särndal, C.-E. (1992). Methods for Estimating the Precision of Survey Estimates When Imputation Has Been Used. Survey Methodology, 18, 241–252.

Song, P.X.-K. (2007). Correlated Data Analysis: Modeling, Analytics, and Applications. Springer Series in Statistics. New York: Springer.

Thomas, E.M. (2004) Aggressive Behaviour Outcomes for Young Children: Change in Parenting Environment Predicts Change in Behaviour. Children and Youth Research Paper Series. Statistics Canada. Catalogue number 89-599-MIE.

Received May 2010
Revised February 2011