University Library

MINERVA ACCESS

A gateway to Melbourne's research publications

Minerva Access is the Institutional Repository of The University of Melbourne

Author/s:
De Silva, AP;De Livera, AM;Lee, KJ;Moreno-Betancur, M;Simpson, JA

Title:
Multiple imputation methods for handling missing values in longitudinal studies with sampling weights: Comparison of methods implemented in Stata

Date:
2021-02-01

Citation:
De Silva, A. P., De Livera, A. M., Lee, K. J., Moreno-Betancur, M. & Simpson, J. A. (2021). Multiple imputation methods for handling missing values in longitudinal studies with sampling weights: Comparison of methods implemented in Stata. BIOMETRICAL JOURNAL, 63 (2), pp.354-371. https://doi.org/10.1002/bimj.201900360.

Persistent Link:
https://hdl.handle.net/11343/276483

Check for updates

Multiple imputation methods for handling missing values in longitudinal studies with sampling weights: comparison of methods implemented in Stata

Running title: Multiple imputation for longitudinal data with sampling weights

Anurika P. De Silva*[,1] , Alysha M. De Livera[1], Katherine J. Lee[2,3], Margarita Moreno-Betancur[1,2,3#], Julie A. Simpson[1#]

[#]Joint last authors

[1]Centre for Epidemiology and Biostatistics, Melbourne School of Population and Global Health, University of Melbourne, Melbourne, Victoria, Australia
[2]Clinical Epidemiology and Biostatistics Unit, Murdoch Children's Research Institute, Royal Children's Hospital, Melbourne, Victoria, Australia
[3]Department of Paediatrics, University of Melbourne, Melbourne, Victoria, Australia

Many analyses of longitudinal cohorts require incorporating sampling weights to account for unequal sampling probabilities of participants, as well as the use of multiple imputation (MI) for dealing with missing data. However, there is no guidance on how MI and sampling weights should be implemented together. We simulated a target population based on the Australian Bureau of Statistics Estimated Resident Population and drew 1,000 random samples dependent on three design variables to mimic the Longitudinal Study of Australian Children (LSAC). The target analysis was the weighted prevalence of overweight/obesity over childhood. We evaluated the performance of several MI approaches available in Stata, based on multivariate normal imputation (MVNI), fully conditional specification (FCS) and two-fold FCS: a weighted imputation model, imputing missing data separately for each quintile sampling weight grouping, including the design stratum indicator in the imputation model, and using sampling weights as a covariate in the imputation model. Approaches based on available cases and inverse probability weighting (IPW), with time-varying weights, were also compared. We observed severe issues of convergence with FCS and two-fold FCS. All MVNI-based approaches performed similarly, producing minimal bias and nominal coverage, except for when imputation was conducted separately for each quintile sampling weight group. IPW performed equally as well as MVNI-based approaches in terms of bias, however, was less precise. In similar longitudinal studies, we recommend using MVNI with the design stratum as a covariate in the imputation model. If this is unknown, including the sampling weight as a covariate is an appropriate alternative.

Key words: Inverse probability weighting; Longitudinal data; Missing data; Multiple imputation; Sampling weight

1 Introduction

Missing data occur frequently in longitudinal cohorts as such studies require the participation of respondents at multiple waves of data collection (Karahalios, Baglietto, Carlin, English, & Simpson, 2012). Multiple imputation (MI) is one approach for handling missing data in longitudinal studies and that has gained popularity in recent times (Hayati Rezvan, Lee, & Simpson, 2015). MI is a two-stage process, where, in the first stage, the incomplete dataset is replicated multiple times, and the missing values are filled by making draws from their posterior distribution given the observed data, estimated from a suitable imputation model. In the second stage, the analysis of interest is performed on each of the imputed datasets, and the parameter estimates are combined using Rubin's rules (Rubin, 1987). Many studies have found that, to ensure that MI gives unbiased results, it is important for the imputation model and the target analysis to be compatible, i.e. a model exists for the joint distribution of the outcome (possibly incomplete) and (other) incomplete variables for which the conditional

* Corresponding author: Centre for Epidemiology and Biostatistics, Melbourne School of Population and Global Health, University of Melbourne, Parkville Campus, 207 Bouverie Street, Melbourne, Victoria 3010, Australia
e-mail: anurika.de@unimelb.edu.au, Phone: +61-383-447-846

This is the author manuscript accepted for publication and has undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process, which may lead to differences between this version and the Version of Record. Please cite this article as doi: 10.1002/bimj.201900360.

This article is protected by copyright. All rights reserved.

Author Manuscript

distributions coincide with the analysis and imputation models (Bartlett & Morris, 2015; Seaman, Bartlett, & White, 2012).

The use of sampling weights is common in longitudinal study designs and adds further complexity to the process of handling missing data. Unequal sampling probabilities are used in population-based studies to ensure that a sufficient number of individuals with rare characteristics or from a minority group are sampled (Seaman & White, 2013). The resulting sample, compared to a random sample drawn from the population, will be less representative of the population. Therefore, when the aim is to perform inferences about the population, sampling weights, defined as the inverse of the probability of being sampled, are applied in the analysis to rebalance the study sample to being representative of the population. This method is known as inverse probability weighting (IPW).

There has been some suggestion that it is important to accommodate sampling weights in the imputation stage of MI (Carpenter & Kenward, 2013; Reiter, Raghunathan, & Kinney, 2006; Seaman & White, 2013; Seaman, White, Copas, & Li, 2012). Indeed, not incorporating the sampling weights into the imputation model may produce biased results due to the incompatibility between the imputation and target analysis models (Reiter et al., 2006; Zhou, Elliott, & Raghunathan, 2016). Yet, guidance on the use of MI for handling missing data in studies with sampling weights is limited. In our personal experience, sampling weights are often ignored in the imputation stage of MI for simplicity.

A number of methods have been proposed in the statistical literature to incorporate complex study design characteristics in the imputation model. A simple method for incorporating the sampling design information is to include the key design variables (e.g. residential area, state or sex) used for stratification in the imputation model (Carpenter & Kenward, 2013) or a stratum indicator (defined on the basis of all possible combinations of values for the design variables) as covariates in the imputation model. A simulation study by Reiter et al. (2006) showed that the latter produced less biased estimates and improved confidence interval coverage compared to disregarding the sampling design information in the imputation model. Alternatively, the sampling weights can be included as a covariate in the imputation model (Carpenter & Kenward, 2013), either in their original form or transformed using the inverse, log transformation or another transformation. A multilevel approach for MI using a random-effects imputation model where the strata used to set the weights are used as the level 2 units was discussed by Carpenter & Kenward (2013) for dealing with missing data in a single variable. Another method suggested is to use a weighted imputation model to reflect the weighted analysis model. However, it has been argued that a weighted imputation model will result in a weighted likelihood which does not relate to a probability model, hence the uncertainty in the imputations may not be reflected accurately (Carpenter, 2011).

Instead of using MI, IPW can also be used to accommodate unequal probability sampling and missing data simultaneously. In addition to weighting the analysis by the sampling weights, the method can be extended to handle missing data by multiplying these by the probability of being a complete case (Hofler, Pfister, Lieb, & Wittchen, 2005), in effect rebalancing the complete cases to be representative of the study sample (Seaman, White, et al., 2012).

A comprehensive evaluation of the methods proposed above, particularly in a longitudinal study setting, has not been conducted. Therefore, the aim of this paper was to evaluate the performance of existing methods for conducting MI in the presence of sampling weights in multivariable missingness settings such as those encountered in longitudinal studies. Specifically, we consider different approaches for accommodating sampling weights or design variables at the imputation stage which can be implemented in the `Stata` statistical software, with the aim of providing guidance to applied researchers who have those approaches available.

We report the findings of a case study from the Longitudinal Study of Australian Children (LSAC) (Australian Institute of Family Studies, 2013) and a simulation study that was designed based on LSAC, where the analysis of interest aimed to estimate the prevalence of overweight/obesity among Australian children over time. The broad MI methods investigated were; multivariate normal imputation (MVNI), fully conditional specification (FCS) and two-fold fully conditional specification (two-fold FCS). MVNI uses a joint imputation model for the vector of all incomplete variables under the assumption of multivariate normality (Schafer, 1997). FCS imputes missing values by fitting univariate imputation models for each incomplete variable including other complete and incomplete variables as predictors (Raghunathan, Lepkowsi, Van Hoewyk, & Solenberger, 2001; Van Buuren, Brand, Groothuis-Oudshoorn, & Rubin, 2006). When handling incomplete longitudinal data, standard implementations of MVNI and FCS can be used by treating the repeated measurements of the same variable at different time points as distinct variables, sometimes referred to as the "Just Another Variable" approach

This article is protected by copyright. All rights reserved.

(Welch, Petersen, et al., 2014). The two-fold FCS algorithm, a version of the standard FCS developed specifically for longitudinal studies, uses variables from only current and adjacent time points as predictors in the imputation, taking into consideration the longitudinal structure of the data to reduce the number of predictors (Nevalainen, Kenward, & Virtanen, 2009; Welch, Petersen, et al., 2014).

Section 2 describes the motivating example, the simulation study, the missing data methods investigated in this paper and the case study. The results and the discussion are the object of Sections 3 and 4 respectively.

2 Methods

2.1 Motivating example: Longitudinal Study of Australian Children

The Longitudinal Study of Australia Children (LSAC) is Australia's first national longitudinal study of children, and was established to explore the social, economic and cultural impact on child and adolescent development (Australian Institute of Family Studies, 2013). In 2004, children were recruited into two cohorts; the infant cohort (B) and the child cohort (K) (see Table S1 in the online Supplementary Material). The first six waves of data collection were available for the current study. Details of the complex sampling design used by LSAC are published elsewhere (Soloff & Johnstone, 2005), but briefly, LSAC used a two-stage clustered sampling design, first selecting postcodes and then children, and stratification was used when selecting postcodes and children to ensure proportional representation for states/territories and urban/rural areas. For simplicity of exposition, the sampling design features considered in the current investigation were child's sex, area (rural or urban) and state of residence (see below for more details).

2.2 Epidemiological analysis of interest

Childhood obesity is a growing problem in most developed countries, with many obese children becoming obese adults (Ebbeling, Pawlak, & Ludwig, 2002; Halfon, Larson, Lu, Tullis, & Russ, 2014; Kiess et al., 2001). The motivating example for our simulation study was an analysis aiming to explore the trends in the prevalence of childhood overweight/obesity in Australia.

Variables of interest. Child's weight status measured at each wave was used to estimate the prevalence of overweight/obesity. In LSAC, child's body mass index (BMI) was measured at each wave, and categorised as underweight III, underweight II, underweight I, normal weight, overweight, and obese, according to International Obesity Task Force age and sex specifications (Cole, Bellizzi, Flegal, & Dietz, 2000; Cole, Flegal, Nicholls, & Jackson, 2007). For simplicity, we dichotomised child's weight status as; overweight/obese, or not. Descriptions of other variables used in this study are provided in Table 1.

<Table 1>

Target analysis. Estimation of the prevalence of overweight/obesity at each wave in the population accounting for the sampling weights using an IPW approach.

2.3 Simulation of the target population

A large target population of 1,000,000 children was simulated to mimic the target population in LSAC (Soloff et al., 2005).

Three design variables, area, state, child's sex which were later used for sample selection, were simulated as follows:

i. Individuals were simulated to be from one of six states ($state_i$) in proportions equal to those provided by the Australian Bureau of Statistics (ABS) 2002 Estimated Resident Population estimates. For simplicity, the Northern Territory and Australian Capital Territory were not considered.

ii. Individual's type of area was simulated as rural or urban ($area_i$), based on the logistic regression model:

$$logit\,\{\mathrm{P}(area_i = 1)\} = \beta_0 + \sum_{a=1}^{5} \beta_{1a}[state_i = a] \quad (1)$$

This article is protected by copyright. All rights reserved.

Author Manuscript

where a = 0 (New South Wales – reference category), 1 (Victoria), 2 (Queensland), 3 (South Australia), 4 (Western Australia), and 5 (Tasmania)

iii. Child's sex (sex_i) was generated by randomly simulating individuals to be male with a 50% probability.

A detailed description of the simulation procedure for the remaining variables, child's ethnicity, age, birthweight, mother's ethnicity, education, BMI at wave 1, mother speaking a language other than English at home, maternal smoking during pregnancy, maternal depression over time, father's BMI, and family socio-economic status z-score is provided in the online Supplementary Material. In summary, data were generated based on the causal diagrams in Figure 1, for each child $i = 1,\ldots,N$, where $N = 1{,}000{,}000$ for waves $j = 1,\ldots,6$.

After generating the time-independent variables (see Figure 1A), the time-dependent variables (see Figure 1B) maternal depression ($mdepression_{ij}$) and the outcome variable of interest, child's weight status ($wtstat_{ij}$) were simulated. The latter was simulated as follows.

i. The variable child's weight status at wave 1 (*wtstati,1*) was randomly generated from a logistic regression model depending on mother's characteristics at baseline (education, language, smoking during pregnancy, BMI), father's BMI at baseline, family socio-economic status z-score, area of residence, state, and current age, sex, ethnicity and birthweight of the child.

ii. The variables for child's weight status at waves j=2,…,6 (*wtstati,j*) were randomly generated from a logistic regression model depending on child's weight status and maternal depression at the previous wave, in addition to the above baseline covariates.

We generated data such that the proportions of overweight/obese children at waves $1,\ldots,6$ were 0.287, 0.250, 0.241, 0.248, 0.263, and 0.286 respectively, to mimic the proportions observed in the LSAC. Similarly, parameter values used in the simulation process were chosen based on the values estimated from analyses of the LSAC child cohort data (see Table S2 in the online Supplementary Material).

<Figure 1>

2.4 Sample selection

A total of 1,000 samples of size 1,000 were drawn from the target population, mimicking the selection of participants into wave 1 of LSAC, based on the logistic regression model:

$$logit\,\{\mathrm{P}(S_i = 1)\} = \Delta_0 + \Delta_1[sex_i = 1] + \Delta_2[area_i = 1] + \sum_{a=1}^{5}\Delta_{3a}[state_i = a] \qquad (2)$$

where S_i is an indicator of selection into the sample, i.e. $S_i = 1$ if an individual was chosen from the population and included in the sample at wave 1, and $S_i = 0$ otherwise.

Parameters in Equation 2 were based on the LSAC data and were chosen such that small subgroups of the population were oversampled (e.g. the probability of rural children being chosen was higher than urban children). See Table 2 for parameter values.

<Table 2>

Estimation of sampling weights at wave 1. The theoretical design weights, known from the design of the study, can be used in analyses. In some cases, it is also possible to estimate the weights from the data using the population level data, which may provide more efficiency in analyses than the theoretical design weights. Whether you use model-estimated response probabilities or true response probabilities there will be no difference in bias, but using estimated probabilities will reduce the variance of the estimate because the observed-data sample is better matched to the complete sample (Lumley, 2010; Robins, Rotnitzky, & Zhao, 1995). A sandwich estimator is used to account for the uncertainty in the weights (Robins, Rotnitzky, & Zhao, 1994; Seaman, White, et al., 2012). In our study, for each of the 1,000 simulated samples, sampling weights were estimated for each child included in the sample at wave 1 by fitting a logistic regression model to the

This article is protected by copyright. All rights reserved.

population of 1,000,000 children deriving predicted selection probabilities given state, area and child's sex and taking the inverse of this probability.

2.5 Introduction of missing data

Missing values were introduced in the time-dependent variables, child's weight status and maternal depression measured across waves 2 to 6. This was done as follows. For each of the 1,000 samples, and at each wave, some children were set as lost to follow-up, and others as having intermittent missingness (i.e. not attending a specific wave, however, coming back at a later wave) under a missing completely at random (MCAR) or missing at random (MAR) mechanism.

For MAR missing data mechanism, it was assumed that the probability of missingness (either due to loss to follow-up or intermittent missingness), followed a logistic regression model, including child's ethnicity ($ethnicity_i$), maternal education ($meducation_i$), language spoken by the mother at home ($mlanguage_i$), family socio-economic status (ses_i), maternal depression ($mdepression_{j-1}$), and child's weight status ($wtstat_{j-1}$) measured at the previous wave. The models used to generate missing values were:

Model A: loss to follow-up

Model A was used to introduce monotone missingness, such that, if the measurement at wave j was assigned to missing using model A, then the individual had measurements missing for all subsequent waves $j+1, \dots, 6$.

$$logit\,[\mathrm{P}(R_{i2}=1)] = \Pi_{0,2} + \Pi_1[ethnicity_i = 1] + \Pi_2[meducation_i = 1] + \Pi_3[mlanguage_i = 1] + \Pi_4 ses_i + \Pi_5[mdepression_{i1} = 1] + \Pi_6[wtstat_{i1} = 1] \quad (3a)$$

$$\mathrm{P}\left(R_{ij}=1 | R_{ij-1}=1\right) = 1; 3 \le j \le 6$$

$$logit\,[\mathrm{P}\left(R_{ij}=1 | R_{ij-1}=0\right)] = \Pi_{0,j} + \Pi_1[ethnicity_i = 1] + \Pi_2[meducation_i = 1] + \Pi_3[mlanguage_i = 1] + \Pi_4 ses_i + \Pi_5[mdepression_{ij-1} = 1] + \Pi_6[wtstat_{ij-1} = 1]; 3 \le j \le 6 \quad (3b)$$

where R_{ij} is an indicator variable for missingness. Weight status and maternal depression were assigned to missing for individual *i* at wave *j* if $R_{ij} = 1$.

Model B: intermittent missingness between waves j-1 and j

Model B was then applied to the individuals who were not specified as lost to follow-up using model A.

$$logit\,[\mathrm{P}\left(R_{ij}=1\right)] = \Psi_{0,j} + \Psi_1[ethnicity_i = 1] + \Psi_2[meducation_i = 1] + \Psi_3[mlanguage_i = 1] + \Psi_4 ses_i + \Psi_5[mdepression_{ij-1} = 1] + \Psi_6[wtstat_{ij-1} = 1];\ 2 \le j \le 6 \quad (4)$$

The parameters of these models were chosen to mimic the missingness patterns in LSAC to represent a weak MAR scenario, and a further strong MAR scenario was explored by doubling the log of the odds ratios used in the weak MAR scenario (see Table 3 for parameter values).

For each mechanism (MCAR or the two MAR mechanisms), the overall missingness proportion (all waves combined) was set at 35% or 65%, representing realistic and extreme missing data scenarios respectively. The required proportions (see Figure S1 in the online Supplementary Material) of missing data at each wave to achieve these overall proportions were obtained by choosing the intercepts for the logistic regression models, $\Pi_{0,j}$ (Equations 3a and 3b) and $\Psi_{0,j}$ (Equation 4) through iteration.

<Table 3>

2.6 Methods for handling missing data

We considered 3 MI methods (MVNI, FCS and two-fold FCS) to multiply impute missing values in child's weight status and maternal depression at waves 2 to 6. Within each of these, we evaluated 5 approaches to incorporate sampling design information within the imputation model as described in Table 4.

This article is protected by copyright. All rights reserved.

Author Manuscript

In the first approach, the sampling weights were excluded from the imputation model. The second approach included the design stratum indicator (derived for this simulation study from all 24 possible combinations of state, area and child's sex) as a covariate (with a fixed effect) in the imputation model.

The rest of the methods explicitly include the sampling weights: including sampling weights (or inverse sampling weights or both) as covariates in the imputation model; grouping sampling weights into 5 strata based on quintile values and imputing missing data separately in each stratum; and accommodating the sampling weights using a weighted imputation model.

We were not able to consider the methods, MVNI with a weighted imputation model, two-fold FCS with a weighted imputation model and imputation separately in quintile sampling weight groupings because these methods are not currently implemented in the statistical software environment used for this simulation study.

All three MI methods treat child's weight status and maternal depression at each wave as distinct variables. FCS and two-fold FCS impute missing data using logistic regression imputation models. FCS includes data from all waves in the imputation, while the two-fold FCS algorithm only uses information from the current and immediately adjacent time points (Welch, Bartlett, & Peterson, 2014). The `Stata` implementation of MVNI fits a joint imputation model assuming multivariate normality, which results in non-binary imputed values for child's weight status and maternal depression. Therefore, we used adaptive rounding to round the non-integer imputed values as either 0 or 1, as recommended by Bernaards, Belin & Schafer (2007) and Lee & Carlin (2010).

The variables mother's age at child birth, ethnicity, education, mother speaking a language other than English at home, maternal smoking during pregnancy, maternal depression at wave 1, mother's and father's BMI at wave 1, family socio-economic status z-score, child's age at wave 1, ethnicity, birthweight, and weight status at wave 1 were included as auxiliary variables in all the imputation models. The three design variables were not explicitly included in the imputation models. A total of 35 or 65 imputations were performed based on the percentage of missingness, as recommended by White, Royston & Wood (2011). After examining convergence diagnostics, a burn-in of 100 iterations and 100 between-imputations were used for the MVNI approaches, a burn in of 10 iterations were used for the FCS approaches, and a time window width of 1 was used for the two-fold FCS approaches with 10 among-time iterations and 5 within-time iterations.

In addition to these MI methods, we evaluated IPW with longitudinal weights, which only included children with observed data at each wave but incorporated the probability of missingness at the given wave in the weighting (see Figure S2 in the online Supplementary Material). Specifically, for each wave, and each individual with observed data at that wave, the longitudinal weight assigned in the target IPW analysis was obtained by multiplying the estimated (wave 1) sampling weight by the inverse of the probability of response at that wave which was estimated from a logistic regression model with the following predictors: mother's education, ethnicity, mother speaking a language other than English at home, family socio-economic status z-score, and maternal depression and child's weight status at each wave.

We also conducted an available case analysis which, at a given wave, excluded children with incomplete body weight status data at that wave and used IPW with the wave 1 sampling weights to estimate the wave-specific prevalence.

<Table 4>

2.7 Performance measures for evaluating different methods

The population proportions of overweight/obesity at waves $j = 2, \ldots, 6$, (0.250, 0.241, 0.248, 0.263, and 0.286), were the true parameters of interest.

We compared the performance of the different MI methods, IPW and available case analysis, in terms of absolute bias (estimated as the difference between the true value and the average of the estimate calculated from the 1,000 samples) and the relative bias (bias relative to true parameter value). The empirical standard error (SE) (square root of variance of 1,000 estimates) and average model-based SE (average of SEs of 1,000 estimates) were computed and these estimates were compared using the relative bias in the SE (estimated as the difference between model-based and empirical SE relative to the empirical SE). The coverage probability of the 95% confidence interval (proportion of simulated datasets in which the true parameter value was contained in the estimated 95% confidence interval) and the mean squared error (MSE), which is a combined measure of bias

This article is protected by copyright. All rights reserved.

and efficiency (Burton, Altman, Royston, & Holder, 2006), are also reported. The Monte Carlo SEs for the estimates (defined as the standard deviations of the estimates over repeated simulations) were used to assess the variation in estimated performance parameters (e.g. bias) across the simulations (White, 2010). For the chosen number of 1,000 simulations, the SE of the estimated coverage was 0.69% for a true coverage of 95%. Therefore, the estimated coverage should lie within the range of 93.6% and 96.4% (with 95% probability) (Morris, White, & Crowther, 2019). (The Monte Carlo SEs are reported in Tables S3-S7 in the supplementary material).

Source code to reproduce the results of the simulation study is available as Supporting Information on the journal's web page (http://onlinelibrary.wiley.com/doi/xxx/suppinfo).

2.8 Case Study

In addition to the simulation study, we provide an empirical comparison of the methods discussed above (excluding FCS and two-fold FCS due to the convergence issues observed with the simulation study, as discussed later) to handle missing longitudinal data with sampling weights using the data from the child cohort of LSAC. As explained previously, for simplicity we used a dichotomised measure (as overweight/obese or not) of the 6-category child's weight status measurement in the current study. For the case study example, all 7 states in Australia were considered (including the Northern Territory and Australian Capital Territory which were excluded from the simulation study) and the sampling weights used in the case study (in the target analysis and imputation models) were obtained from LSAC. A total of 65 imputations were performed to account for the missing data in the time-dependent outcome child's weight status and the auxiliary variables included in the imputation models.

`Stata 13.1` statistical software (Stata Corporation, 2013) was used for all analyses.

3 Results

3.1 Results from the simulation study

Issues of convergence. For all of the imputation models, FCS was problematic in terms of convergence of the procedure, i.e. the program crashed in a large proportion of the simulated datasets and did not obtain imputed values. In particular FCS applied separately in the quintile groupings of the sampling weights did not converge in any of the six evaluated scenarios. As expected, the frequency of non-convergence under the FCS methods increased in the presence of a larger proportion of missing data.

In all versions considered, the two-fold FCS approach faced issues of convergence, similar to FCS. However, as expected, minimal to substantially lower frequencies of non-convergence were observed across scenarios for two-fold FCS compared with FCS, due to two-fold FCS using information from only current and immediately adjacent time points for imputation, i.e. imputation models have less predictors. Due to the convergence issues faced by FCS and the two-fold FCS, only the performance of MVNI will be discussed from this point forward. For completeness, performance measures corresponding to the scenarios where FCS and the two-fold FCS converged are presented in the online Supplementary Material.

The issues of convergence were minimal under MVNI, with only the approach of imputing separately in the quintile groupings of the sampling weights facing issues of convergence in some (0.1%-5.5%) of the datasets in the presence of 65% missing data. Using the quintile approach resulted in biased estimates with large standard errors and a poor coverage probability (as low as 89.21%; see Table S3 in the online Supplementary Material). This approach will be excluded from discussion henceforth, however, results from this method are shown in Figures 2-4 along with the other MVNI-based methods.

Bias. All remaining MVNI-based methods performed similarly, producing minimal bias for the scenarios with 35% missingness (see Figure 2). Slight differences in the bias were observed for the missingness percentage of 65%, especially for the latter time points. As expected, excluding the sampling weights from the imputation model produced relatively more bias compared to the other imputation methods. For the strong MAR scenario with 65% missingness, MVNI with the design stratum indicator included in the imputation model resulted in the

This article is protected by copyright. All rights reserved.

smallest relative bias of 0.24%, while the imputation model without the sampling weights resulted in the largest relative bias of 2.13% (see Table S3 in the online Supplementary Material).

When compared with IPW, all MVNI-based methods performed equally well or better in terms of bias. The relative bias for IPW ranged from 0.04% to 3.97%, while the relative bias for all MVNI-based methods ranged from 0.005% to 2.46%. Available case analyses resulted in substantially larger bias in most scenarios compared to MVNI and IPW and resulted in the largest bias of 14.53% for 65% missingness under the MAR strong scenario (see Table S7 in the online Supplementary Material).

Precision. There was not much of a difference between the MVNI-based methods in terms of precision. All the methods resulted in small model-based and empirical standard errors ranging from 0.02 to 0.03. All MVNI-based methods showed some gain in precision compared to IPW (model-based and empirical standard errors ranging from 0.02-0.05) and available case analysis for a few scenarios (model-based and empirical standard errors ranging from 0.02-0.03). Available case analysis, IPW and MVNI without sampling weights resulted in a higher relative bias in the standard errors compared to the other MVNI-based methods which resulted in very similar model-based and empirical standard errors.

Mean squared error. We observed expected patterns for MSE given the bias and precision findings stated above with both IPW and available case analysis performing poorly compared to MVNI, especially with larger proportions of missing data and a stronger MAR scenario (see Figure 3). Also similar to what we observed with the precision, there was a similar MSE with all of the MVNI-based methods.

Coverage. The coverage was within 93.6% and 96.4% for the nominal level of 95% (expected range for coverage based on 1,000 simulations) for all the MVNI-based methods and IPW, and there was similar coverage among the MVNI-based approaches. However, available case analysis resulted in poor coverage probabilities in most scenarios, with a coverage of 70.5% when there was 65% incomplete data under the strong MAR scenario.

<Figure 2>

<Figure 3>

<Figure 4>

3.2 Results from the case study

Consistent with the findings of our simulation study, all MVNI-based approaches produced similar estimates and 95% confidence intervals (see Figure 5). As for the simulation study, the method of imputing separately in quintile sampling weight groupings did not converge. Available case analysis and IPW produced slightly different estimates and slightly larger confidence intervals compared to the MVNI-based approaches and these differences became more evident during the latter waves (see Table S8 in the online Supplementary Material).

<Figure 5>

4 Discussion

We compared a number of different imputation methods which can be implemented in the `Stata` statistical software within the MI framework, as well as IPW and available case analysis, for handling missing longitudinal data in a study where the analysis required incorporating sampling weights. Our example involved handling missing data in child's weight status, recorded at six different ages, where the analysis of interest aimed to estimate the population prevalence of overweight/obesity using sampling weights to account for the unequal probability sampling in the study design. We considered two missing data proportions (35% and 65%) and three missing data scenarios (MCAR, MAR weak and MAR strong).

Similar to the findings of Welch, Petersen, et al. (2014), Kalaycioglu, Copas, King & Omar (2015), and De Silva, Moreno-Betancur, De Livera, Lee & Simpson (2019) we observed severe issues of non-convergence (in up to 100% of the simulated datasets) when FCS was used to impute child's weight status for five waves of data. This is most likely due to the many explanatory variables included in the imputation model which may have

This article is protected by copyright. All rights reserved.

Author Manuscript

caused numerical issues because of model overfitting and/or issues of collinearity due to a number of correlated repeated measures. We also observed the problem of perfect prediction and using augmented logistic regression imputation did not lower the rate of non-convergence, similar to our previous findings in De Silva et al. (2019). We compared the standard FCS method in `Stata` and the '`mice`' package in `R` as all the studies cited above which reported issues of convergence, implemented the standard FCS method in `Stata`. We observed no issues of convergence in `R`, however, the method produced significantly more biased and less precise estimates compared to the corresponding MVNI-based approach. Therefore, this did not change the recommendations made in our study. Consistent with previous studies (Kalaycioglu et al., 2015; Welch, Petersen, et al., 2014) we observed reductions in the rate of non-convergence for two-fold FCS compared to standard FCS (especially for high missing data proportions (65%) and stronger MAR scenarios), however, similar to the findings of De Silva et al. (2019) it did not reduce the amount of non-convergence to a negligible amount. Therefore, neither of the FCS methods were suitable for most scenarios considered in our simulation study.

Consistent with the findings of the studies in the statistical literature evaluating the performance of MVNI when imputing missing values in binary data (Lee & Carlin, 2010; Schafer, 1997) we observed all the MVNI-based approaches performed well in terms of bias and precision. MVNI faced only minimal issues of non-convergence, which occurred when used to impute missing values separately in quintile sampling weight groupings in settings with higher proportions of missingness, which led to a very small number of individuals with complete data in some cases. Using tertile or quartile sampling weight groupings may reduce convergence problems, or alternatively the quintile groupings of the sampling weights could be included as a fixed effect in the imputation model.

We observed that the performance of all MVNI-based methods (with the exception of imputing separately in quintile sampling weight groupings) was similar (i.e. negligible bias and nominal coverage) in the presence of 35% missing data, whereas for 65% missing data a small amount of bias was observed for some imputation models. This is consistent with the findings of a study by Reiter et al. (Reiter et al., 2006), which analysed National Health and Nutrition Examination Surveys (1999-2002), and observed that when handling a smaller percentage of missing values (15% missing data) accounting for study design features in the imputation model may not have a substantial impact on the parameter estimates.

As emphasised in the statistical literature excluding sampling weights from the imputation model leads to issues of compatibility between the imputation and analysis models (Reiter et al., 2006; Zhou et al., 2016), which can lead to biased results (von Hippel, 2009; White et al., 2011). The literature also states that if the variables used to derive the sampling weights are known then the inclusion of these variables in the imputation model may address the issues of incompatibility (Zhou et al., 2016). Including the design variables child's sex, area and state of residence (as covariates or design stratum indicator), performed better than the imputation model without the design variables, however, for the latter the relative bias was surprisingly still small in our simulation setting, 2.46% for 65% missing data. Further exploration, excluding auxiliary variables (associated with the design variables) from the imputation model produced a relative bias of up to 7.65% (see online Supplementary Material). This suggests that simply incorporating auxiliary variables which act as proxies for the sampling design information when these design variables are not available to the analyst/data user, may also reduce bias due to incompatibility (Zhou et al., 2016).

Our analysis of interest aimed to estimate the prevalence of overweight/obesity at each wave of data collection, hence no other variables were included in the analysis model. For analysis models with multiple variables, it has been shown that interactions between the sampling weight covariate and all the variables in the analysis model need to be included in the imputation model to reduce bias (Carpenter & Kenward, 2013; Kim, Brick, Fuller, & Kalton, 2006; Seaman, White, et al., 2012; Zhou et al., 2016). These additional interaction terms may lead to non-convergence of the MI procedures, especially in the longitudinal setting where there are often missing data in many time dependent variables. One proposed solution is to use a random-effects imputation model with quantiles of sampling weights defining level 2 units, recommended by Carpenter and Kenward (Carpenter & Kenward, 2013). The performance of this method has been evaluated in a cross-sectional case study setting and implementation with longitudinal data would need software allowing for a joint multivariate mixed model with more than 2 levels, which is not available yet in standard statistical software to our knowledge.

The design stratum indicator in our study was derived from three design variables, child's sex, area and state of residence, resulting in 24 strata, and thus 23 dummy variables being included in the imputation model. The performance of this method may not be as good in studies with a larger number of strata, thereby, requiring

This article is protected by copyright. All rights reserved.

aggregation of strata with similar characteristics (Reiter et al., 2006) which could lead to further issues of non-convergence.

As expected, the available case analysis produced biased estimates and very low coverage compared to the MVNI-based approaches and IPW. While IPW uses only the data from the available cases in each wave in the target analysis, it weights each individual by the inverse probability of being a respondent in that wave, which corrects the inherent selection bias (Seaman & White, 2013). Consistent with previously published studies which compared MI and IPW for handling missing data (Beunckens, Sotto, & Molenberghs, 2008; Carpenter, Kenward, & Vansteelandt, 2006; Seaman & White, 2013; Vansteelandt, Carpenter, & Kenward, 2010), we observed MI-based methods to be more efficient than IPW. There are a number of reasons for this including the ability to use auxiliary variables in the imputation models which are strong predictors of the missing values (De Silva, Moreno-Betancur, De Livera, Lee, & Simpson, 2017; Graham & Collins, 2012; Karahalios et al., 2013). Even with a correctly specified missingness model IPW may be less efficient than MI with a correctly specified imputation model (Seaman & White, 2013). A number of studies have recommended possible methods for increasing the efficiency of IPW estimators, including methods such as weight stabilisation (Seaman & White, 2013), augmentation of the IPW equations (Vansteelandt et al., 2010) and more generally doubly robust IPW estimators (Carpenter et al., 2006).

Zhou, Elliott & Raghunathan (2016) proposed a two-step semiparametric approach for handling sampling weights and missing data due to item non-response at a single time point. This method can be currently implemented in the `R` statistical software and in the Imputation and Variance Estimation (IVEware) Software in `R`, `Stata`, `SAS` and `SPSS` (Michigan, 2020). While the authors have made suggestions to extend this method to handle unit non-response, an extension for repeated measures data has not yet been proposed and how this method should be extended to the longitudinal setting requires further research. Another approach to handle missing data and sampling weights simultaneously is to use an extension of the fully Bayesian approach for handling missing data (Goldstein, Carpenter, & Browne, 2014). This approach recommended by Goldstein, Carpenter, & Kenward (2018), can be used in certain scenarios. This 'Hybrid Markov chain Monte Carlo algorithm for Bayesian estimation incorporating weights' uses a Bayesian model for handling missing data and the information from the sampling weights are incorporated through a bootstrap. Through simulation studies and a case study Goldstein et al. (2018) showed that this method resulted in small bias and good coverage in most scenarios except in the presence of a large number of strata. The authors recommended MI methods such as those investigated in our simulation study as an alternative in those cases. Unfortunately, the accessibility of this method is limited; it can be implemented in `MATLAB` (Mathworks, 2015) only at present and thus its evaluation was out of the scope of the present paper.

The strength of the current study is that we simulated a large target population, based on National data (Soloff et al., 2005) and the child cohort of LSAC, to ensure our sampling weights and inter-relatedness of the parental and child variables captured the complexity of the LSAC cohort, compared with other investigations which have used a more simplified simulation framework (Goldstein et al., 2018; Reiter et al., 2006). Simulation studies based on real cohort studies are now frequently being used in the statistical literature (De Silva et al., 2017, 2019; Karahalios et al., 2013; Marshall, Altman, Royston, & Holder, 2010), as it allows complex yet realistic associations to be incorporated into the simulated data. We also provide a case study from the LSAC child cohort for an empirical illustration of our observations in the simulation study. One limitation is that we only explored a weighted analysis model which estimated the prevalence of overweight/obesity at each wave as the influence of sampling weights is high in such scenarios (Heeringa, West, & Berglund, 2010). Further research investigating the performance of the methods for incomplete data with sampling weights in the presence of complex longitudinal analysis models is warranted. Caution is required when generalising these results as our simulation study was designed based on a single cohort. Furthermore, the MI methods investigated in our study are only guaranteed to be unbiased when the missing data are MAR as in our study (Carlin, 2015). Further research on the performance of these MI-based approaches for deviations from MAR is important (Hayati Rezvan et al., 2015), however, was beyond the scope of our paper.

To conclude, this simulation study showed that when imputing missing longitudinal data when the target analysis involves sampling weights, there were major convergence issues for both FCS and two-fold FCS. With the exception of imputing separately in quintile sampling weight groupings, MVNI performed well for all imputation models investigated. Due to simplicity, we recommend the use of MVNI with the design stratum indicator (derived from all possible combinations of the design variables) included as a covariate in the imputation model. However, if the design stratum indicator is unknown, sampling weights can be included as a covariate in the imputation model with MVNI. Including sampling weights as a covariate in the imputation

This article is protected by copyright. All rights reserved.

model should, however, be used with caution as this may require the imputation model to include interaction terms between sampling weights and variables in the analysis model.

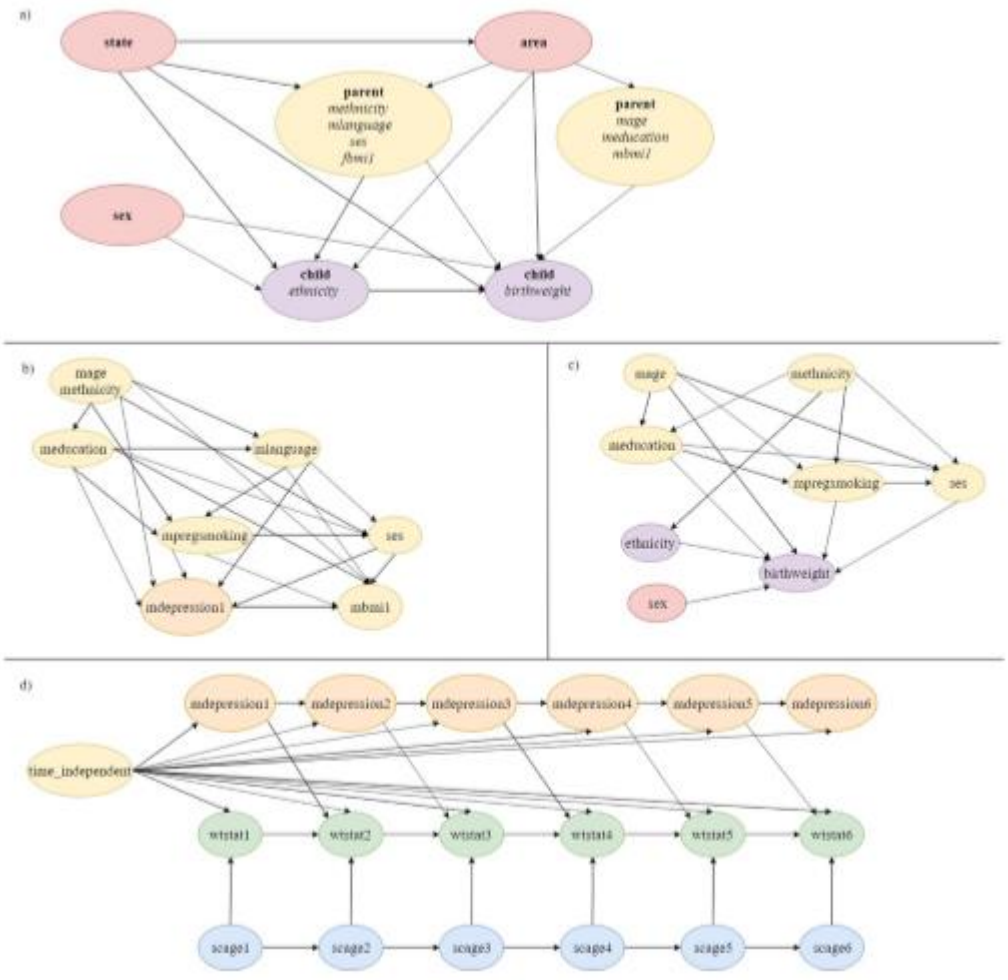
Supporting Information
Additional supporting information is available in Supplementary Material for publication.

Acknowledgements
This work was supported by funding from the National Health and Medical Research Council: a Centre of Research Excellence grant, ID 1035261, awarded to the Victorian Centre of Biostatistics (ViCBiostat); and a Senior Research Fellowship ID 1104975 (JAS) and Career Development Fellowship ID 1053609 (KJL).

Conflict of Interest
The authors have declared no conflict of interest.

Availability of data and materials
The data that support the findings of this study and the source code to reproduce the results are available from the corresponding author upon reasonable request.

Ethics approval and consent to participate
For the simulation study, data were completely simulated, which did not require approval from the ethics committee or consent from participants. The case study example used in this study was based on the child cohort of LSAC which has been provided ethical clearance by the Australian Institute of Family Studies Ethics Committee. The authors are approved users of the LSAC data and were granted access to the data through The University of Melbourne's Organisational Deed of License. Written informed consent was obtained from the caregiver on behalf of each of the study children, as the children were minors at the time of data collection. The signed consent forms are retained by the field agency (Australian Bureau of Statistics).



This article is protected by copyright. All rights reserved.

Figure 1 Simplified casual diagram for the simulation of a) the three design variables; state, area and child's sex and their associations with parent and child variables; b) maternal time-independent variables (for simplicity state and area variables are not shown and mother's age and ethnicity are shown as a single node); c) child variables (for simplicity only maternal variables associated with the child variables are shown); d) time-dependent variables (for simplicity all time-independent variables are shown as a single node).
[state - state of residence; area - area of residence; methnicity - mother's ethnicity; mage - maternal age at child birth; meducation - maternal education; mpregsmoking - maternal smoking during pregnancy; mlanguage - language spoken other than English at home by the mother; mbmi1 - mother's body mass index (BMI) at wave 1; fbmi1 - father's BMI at wave 1; ses - family socio-economic status at wave 1; sex - child's sex; ethnicity - child's ethnicity; birthweight - child's birth weight; mdepression1 - maternal depression at wave 1 (only shown in Figure 1b due to mother's BMI at wave 1 dependency); mdepression1-mdepression6 - maternal depression at waves 1 to 6; wtstat1 – wtstat6 - child's weight status at waves 1 to 6; scage1 – scage6 - child's age at waves 1 to 6].

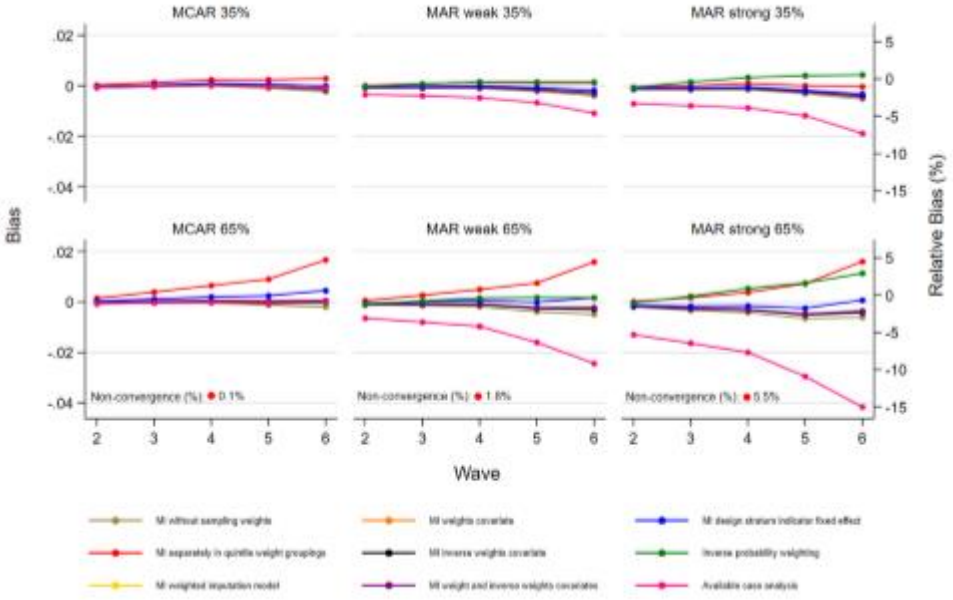


Figure 2 Bias (compared with true prevalence of 0.250, 0.241, 0.248, 0.263, and 0.286 for waves 2 to 6 respectively) and Relative Bias (%) for multiple imputation (MI) without using sampling weights in the imputation model (MI without sampling weights); MI applied separately in the quintile groupings of the sampling weights (MI separately in quintile weight groupings); MI conducted with a weighted imputation model (MI weighted imputation model); MI with sampling weights as a covariate in the imputation model (MI weights covariate); MI with inverse sampling weights as a covariate in the imputation model (MI inverse weights covariate); MI with both weights and inverse weights as covariates in the imputation model (MI weights and inverse weights covariates); MI with the design stratum indicator (derived from all 24 possible combinations of state, area and child's sex) included as a fixed effect in the imputation model (MI design stratum indicator fixed effect) for handling increasing proportions of missing data (35%, 65%) in child's weight status, for the estimated prevalence of overweight/obesity at wave $j = 2, \ldots, 6$ when data are missing completely at random (MCAR), missing at random weak (MAR weak) and missing at random strong (MAR strong).
[a]The results are presented for the MI method multivariate normal imputation (MVNI). The current Stata implementation of MVNI does not allow for a weighted imputation model. [b]Inverse probability weighting (IPW) and available case analysis (ACA) are included for comparison. [c]Percentage non-convergence for each method is shown in each figure. Only methods which faced issues of convergence are listed.

This article is protected by copyright. All rights reserved.

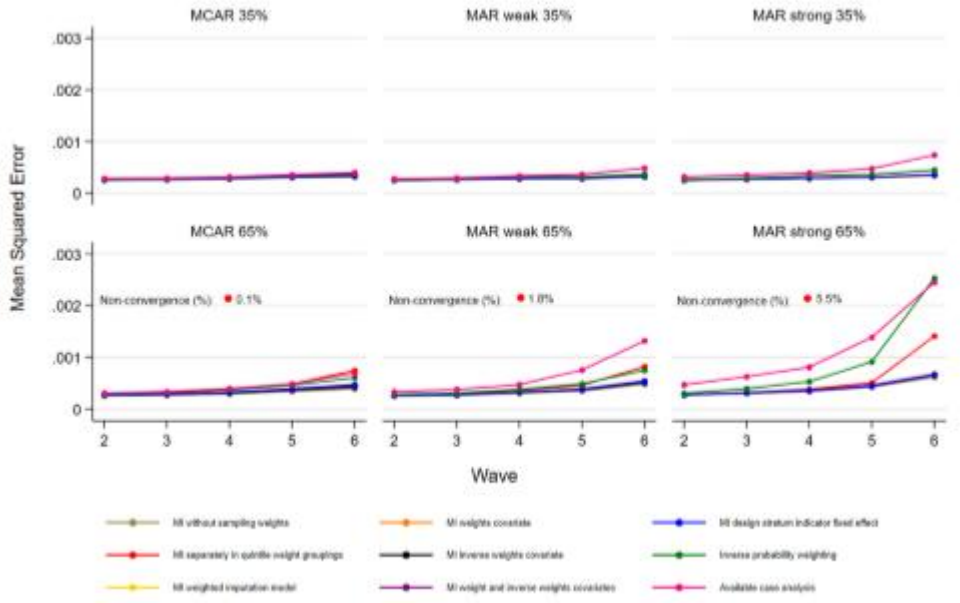


Figure 3 Mean squared error (a combined measure of bias and precision ($Bias^2 + Variance$) for multiple imputation (MI) without using sampling weights in the imputation model (MI without sampling weights); MI applied separately in the quintile groupings of the sampling weights (MI separately in quintile weight groupings); MI conducted with a weighted imputation model (MI weighted imputation model); MI with sampling weights as a covariate in the imputation model (MI weights covariate); MI with inverse sampling weights as a covariate in the imputation model (MI inverse weights covariate); MI with both weights and inverse weights as covariates in the imputation model (MI weights and inverse weights covariates); MI with the design stratum indicator (derived from all 24 possible combinations of state, area and child's sex) included as a fixed effect in the imputation model (MI design stratum indicator fixed effect) for handling increasing proportions of missing data (35%, 65%) in child's weight status, for the estimated prevalence of overweight/obesity at wave $j = 2, \ldots, 6$ when data are missing completely at random (MCAR), missing at random weak (MAR weak) and missing at random strong (MAR strong).
[a]The results are presented for the MI method multivariate normal imputation (MVNI). The current Stata implementation of MVNI does not allow for a weighted imputation model. [b]Inverse probability weighting (IPW) and available case analysis (ACA) are included for comparison. [c]Percentage non-convergence for each method is shown in each figure. Only methods which faced issues of convergence are listed.

Author Man

This article is protected by copyright. All rights reserved.

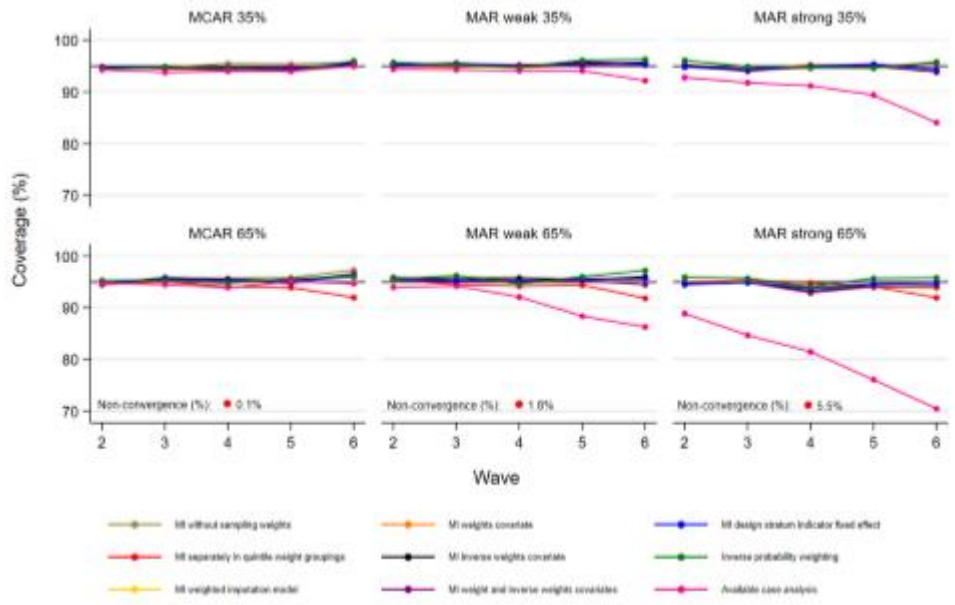


Figure 4 Coverage (proportion of simulated datasets in which the true prevalence of overweight/obesity of 0.250, 0.241, 0.248, 0.263, and 0.286 for waves 2 to 6 respectively was contained in the estimated 95% confidence interval) for multiple imputation (MI) without using sampling weights in the imputation model (MI without sampling weights); MI applied separately in the quintile groupings of the sampling weights (MI separately in quintile weight groupings); MI conducted with a weighted imputation model (MI weighted imputation model); MI with sampling weights as a covariate in the imputation model (MI weights covariate); MI with inverse sampling weights as a covariate in the imputation model (MI inverse weights covariate); MI with both weights and inverse weights as covariates in the imputation model (MI weights and inverse weights covariates); MI with the design stratum indicator (derived from all 24 possible combinations of state, area and child's sex) included as a fixed effect in the imputation model (MI design stratum indicator fixed effect) for handling increasing proportions of missing data (35%, 65%) in child's weight status, for the estimated prevalence of overweight/obesity at wave $j = 2, \ldots, 6$ when data are missing completely at random (MCAR), missing at random weak (MAR weak) and missing at random strong (MAR strong).
[a]The results are presented for the MI method multivariate normal imputation (MVNI). The current Stata implementation of MVNI does not allow for a weighted imputation model. [b]Inverse probability weighting (IPW) and available case analysis (ACA) are included for comparison. [c]Percentage non-convergence for each method is shown in each figure. Only methods which faced issues of convergence are listed.

Author Manuscript

This article is protected by copyright. All rights reserved.

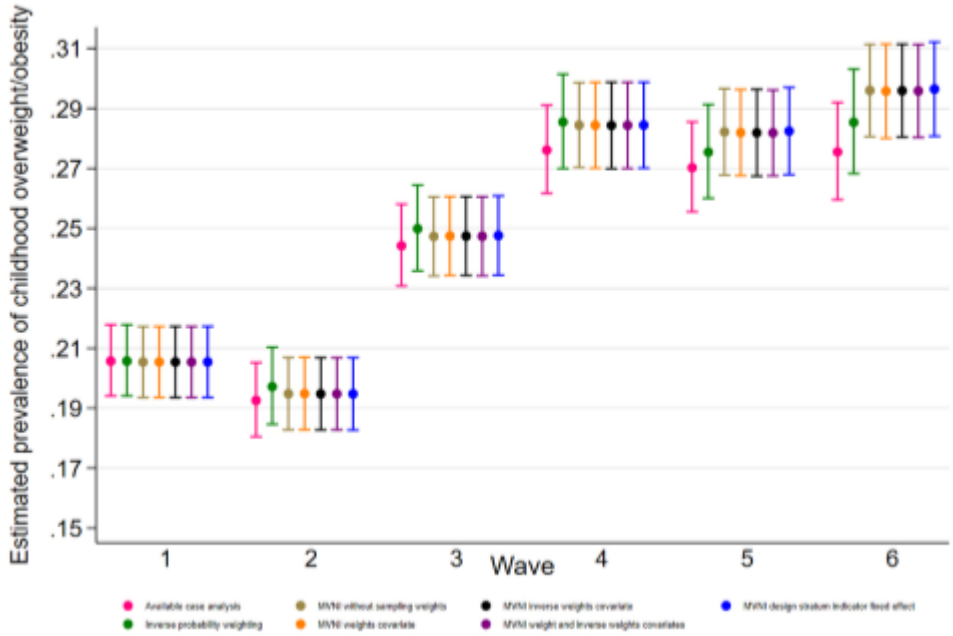


Figure 5 Estimated prevalence of childhood overweight/obesity and 95% confidence intervals for the case study analysis (child cohort of the Longitudinal Study of Australian Children) obtained from a weighted analysis model using an available case analysis, inverse probability weighting, multivariate normal imputation (MVNI) without using sampling weights in the imputation model (MVNI without sampling weights), MVNI with sampling weights as a covariate in the imputation model (MVNI weights covariate), MVNI with inverse sampling weights as a covariate in the imputation model (MVNI inverse weights covariate), MVNI with both weights and inverse weights as covariates in the imputation model (MVNI weights and inverse weights covariates) and MVNI with the design stratum indicator included as a fixed effect in the imputation model (MVNI design stratum indicator fixed effect) for handling missing data in child's weight status measured repeatedly across 6 waves.

[a]Results are not shown for MVNI separately in quintile weight groupings because the imputation model failed to converge and the current Stata implementation of MVNI does not allow for a weighted imputation model.

Author Man

This article is protected by copyright. All rights reserved.

Table 1 Description of variables from the Longitudinal Study of Australian Children used in the simulation study. Notation defined for individual *i* at wave *j*

Variable	Type	Grouping/ Units	Notation
Study child's weight status at wave *j*=1,…,6 (Is the child overweight/obese?)	Binary	0 = No 1 = Yes	$wtstat_{ij}$
Study child's sex	Binary	0 = Female 1 = Male	sex_i
Study child's ethnicity (Indigenous background)	Binary	0 = No 1 =Yes	$ethnicity_i$
Study child's age	Continuous	Months	$scage_{ij}$
Study child's birth weight	Continuous	kilograms	$birthweight_i$
Mother's ethnicity (Indigenous background)	Binary	0 = No 1 =Yes	$methnicity_i$
Maternal education (Mother completed school)	Binary	0 = No 1 =Yes	$meducation_i$
Language spoken other than English at home by the mother	Binary	0 = No 1 =Yes	$mlanguage_i$
Maternal smoking during pregnancy	Binary	0 = No 1 =Yes	$mpregsmoking_i$
Maternal depression at wave j=1,…,6	Binary	0 = No 1 = Yes	$mdepression_{ij}$
Mother's body mass index at wave 1	Continuous	kg/m^2	$mbmi_{i1}$
Father's body mass index at wave 1	Continuous	kg/m^2	$fbmi_{i1}$
Family socio-economic status at wave 1	Continuous	z-score	ses_{i1}
Area	Binary	0 = Rural 1 = Urban	$area_i$
State	Categorical	0 = NSW 1 = VIC 2 = QLD 3 = SA 4 = WA 5 = TAS	$state_i$

Abbreviations: NSW, New South Wales; QLD, Queensland; SA, South Australia; TAS, Tasmania; VIC, Victoria; WA, Western Australia

Table 2 Parameter values used in the logistic regression model for sample selection (Equation 2)

Variable	Parameter value (log Odds Ratio for sample selection)
Sex_{male}	$\Delta_1 = \log(0.95)$
$Area_{urban}$	$\Delta_2 = \log(0.61)$
State	
$State_{New\ South\ Wales}$	$\Delta_{30} = \log(1.00)$
$State_{Victoria}$	$\Delta_{31} = \log(1.08)$
$State_{Queensland}$	$\Delta_{32} = \log(1.32)$
$State_{South\ Australia}$	$\Delta_{33} = \log(4.26)$
$State_{Western\ Australia}$	$\Delta_{34} = \log(2.72)$
$State_{Tasmania}$	$\Delta_{35} = \log(5.21)$

Abbreviations: log, natural logarithm

Author Manuscript

This article is protected by copyright. All rights reserved.

Table 3 Specifications of the parameters in the logistic regression models used to impose missing data in child's weight status and maternal depression under the missing at random (MAR) scenario (Equations 3 and 4)

Variable	log Odds Ratio			
	MAR weak		MAR strong[a]	
	Model A Equation 3[b]	Model B Equation 4[b]	Model A Equation 3[b]	Model B Equation 4[b]
Child's ethnicity	$\Pi_1 = \log(1.97)$	$\Psi_1 = \log(1.88)$	$\Pi_1 = \log(3.88)$	$\Psi_1 = \log(3.53)$
Maternal education	$\Pi_2 = \log(0.82)$	$\Psi_2 = \log(0.87)$	$\Pi_2 = \log(0.67)$	$\Psi_2 = \log(0.76)$
Language spoken by the mother	$\Pi_3 = \log(1.07)$	$\Psi_3 = \log(1.02)$	$\Pi_3 = \log(1.14)$	$\Psi_3 = \log(1.04)$
Family socio-economic status	$\Pi_4 = \log(0.94)$	$\Psi_4 = \log(0.96)$	$\Pi_4 = \log(0.88)$	$\Psi_4 = \log(0.92)$
Maternal depression at wave j-1	$\Pi_5 = \log(1.07)$	$\Psi_5 = \log(1.02)$	$\Pi_5 = \log(1.14)$	$\Psi_5 = \log(1.04)$
Child's weight status at wave j-1	$\Pi_6 = \log(1.09)$	$\Psi_6 = \log(1.05)$	$\Pi_6 = \log(1.19)$	$\Psi_6 = \log(1.10)$

Abbreviations: log, natural logarithm

[a] Odds ratio for MAR strong=square of the Odds ratio for MAR weak

[b] Models A and B represent the logistic regression models used to generate lost to follow-up and intermittent missingness respectively, from waves 2-6 under MAR.

Table 4 Description of methods for handling missing data

Method	Description	Sampling weights for target analysis	Multiple Imputation Methods considered
MI excluding sampling weights	Imputation is conducted without sampling weights included in the imputation model followed by a weighted analysis	Sampling weights at wave 1	MVNI FCS two-fold FCS
MI with design stratum indicator as a covariate in the imputation model	Include the design stratum indicator (derived from all 24 possible combinations of state, area and child's sex) as a covariate (with a fixed effect) in the imputation model followed by a weighted analysis	Sampling weights at wave 1	MVNI FCS two-fold FCS
MI using a weighted imputation model	The imputation model is weighted using sampling weights at wave 1 followed by a weighted analysis	Sampling weights at wave 1	FCS
MI using sampling weights/ inverse weights/ both weights and inverse weights as a covariate(s) in the imputation model	Include sampling weights/ inverse of sampling weights/ both weights and inverse weights at wave 1 as a covariate(s) in the imputation model followed by a weighted analysis	Sampling weights at wave 1	MVNI FCS two-fold FCS
MI applied separately in quintile sampling weight groupings	Imputation conducted separately in the quintile groupings of the sampling weights at wave 1 followed by a weighted analysis	Sampling weights at wave 1	MVNI FCS
Available case analysis	Exclude children with incomplete data at a given wave followed by a weighted analysis	Sampling weights at wave 1	

This article is protected by copyright. All rights reserved.

Inverse probability weighting	Exclude children with incomplete data at a given wave followed by a weighted analysis	Longitudinal sampling weights[a]

Abbreviations: FCS, Fully conditional specification; MI, Multiple imputation; MVNI, Multivariate normal imputation; two-fold FCS, two-fold fully conditional specification.

[a]Inverse probability weighting uses only the data from the available cases in each wave in the target analysis and for each wave, and each individual with observed data at that wave, the longitudinal weight assigned in the target IPW analysis was obtained by multiplying the estimated (wave 1) sampling weight by the inverse of the probability of response at that wave which was estimated from a logistic regression model with the following predictors: mother's education, ethnicity, mother speaking a language other than English at home, family socio-economic status z-score, and maternal depression and child's weight status at each wave.

Author Manuscript

This article is protected by copyright. All rights reserved.

References

Australian Institute of Family Studies. (2013). *The Longitudinal Study of Australian Children: An Australian Government Initiative, Data User Guide*.

Bartlett, J., & Morris, T. P. (2015). Multiple imputation of covariates by substantive model compatible fully conditional specification. *The Stata Journal, 15*(2), 437-456.

Bernaards, C. A., Belin, T. R., & Schafer, J. L. (2007). Robustness of a multivariate normal approximation for imputation of incomplete binary data. *Stat Med, 26*(6), 1368-1382.

Beunckens, C., Sotto, C., & Molenberghs, G. (2008). A simulation study comparing weighted estimating equations with multiple imputation based estimating equations for longitudinal binary data. *Computational Statistics & Data Analysis, 52*(3), 1533-1548.

Burton, A., Altman, D. G., Royston, P., & Holder, R. L. (2006). The design of simulation studies in medical statistics. *Stat Med, 25*(24), 4279-4292.

Carlin, J. B. (2015). Multiple imputation: a perspective and historical overview. In: *Handbook of Missing Data* (eds. G. Molenberghs, G. M. Fitzmaurice, M. G. Kenward, A. A. Tsiatis & G. Verbeke), Boca Raton: Chapman & Hall/CRC Press.

Carpenter, J. (2011). *Multiple imputation with survey weights- a bad idea?* Retrieved from *www.missingdata.org.uk*. Accessed: 5 January 2017.

Carpenter, J., Kenward, M. G., & Vansteelandt, S. (2006). A comparison of multiple imputation and doubly robust estimation for analyses with missing data. *Journal of the Royal Statistical Society Series A (Statistics in Society), 169*(3), 571-584.

Carpenter, J. R., & Kenward, M. G. (2013). *Multiple imputation and its application*: Chichester: Wiley.

Cole, T. J., Bellizzi, M. C., Flegal, K. M., & Dietz, W. H. (2000). Establishing A Standard Definition For Child Overweight And Obesity Worldwide:International Survey. *British Medical Journal, 320*(7244), 1240-1243.

Cole, T. J., Flegal, K. M., Nicholls, D., & Jackson, A. A. (2007). Body mass index cut offs to define thinness in children and adolescents: international survey. *bmj, 335*(7612), 194.

De Silva, A. P., Moreno-Betancur, M., De Livera, A. M., Lee, K. J., & Simpson, J. A. (2017). A comparison of multiple imputation methods for handling missing values in longitudinal data in the presence of a time-varying covariate with a non-linear association with time: a simulation study. *BMC Med Res Methodol, 17*(1), 114-124.

De Silva, A. P., Moreno-Betancur, M., De Livera, A. M., Lee, K. J., & Simpson, J. A. (2019). Multiple imputation methods for handling missing values in a longitudinal categorical variable with restrictions on transitions over time: a simulation study. *BMC Med Res Methodol, 19*(1), 14.

Ebbeling, C. B., Pawlak, D. B., & Ludwig, D. S. (2002). Childhood obesity: public-health crisis, common sense cure. *The Lancet, 360*(9331), 473-482.

Goldstein, H., Carpenter, J., & Browne, W. J. (2014). Fitting multilevel multivariate models with missing data in responses and covariates that may include interactions and non-linear terms. *Journal of the Royal Statistical Society Series A (Statistics in Society), 177*(2), 553-564.

Goldstein, H., Carpenter, J., & Kenward, M. G. (2018). Bayesian models for weighted data with missing values: a bootstrap approach. *Journal of the Royal Statistical Society Series C (Applied Statistics), 67*(4), 1071-1081.

Graham, J. W., & Collins, L. M. (2012). Using modern missing data methods with auxiliary variables to mitigate the effects of attrition on statistical power. In *Missing Data* (pp. 253-275). New York: Springer.

Halfon, N., Larson, K., Lu, M., Tullis, E., & Russ, S. (2014). Lifecourse health development: past, present and future. *Matern Child Health J, 18*(2), 344-365.

Hayati Rezvan, P., Lee, K. J., & Simpson, J. A. (2015). The rise of multiple imputation: a review of the reporting and implementation of the method in medical research. *BMC Med Res Methodology, 15*, 30-43.

Heeringa, S., West, B. T., & Berglund, P. A. (2010). *Applied survey data analysis*: Boca Raton, FL : Chapman & Hall/CRC.

Hofler, M., Pfister, H., Lieb, R., & Wittchen, H. U. (2005). The use of weights to account for non-response and drop-out. *Soc Psychiatry Psychiatr Epidemiol, 40*(4), 291-299. doi:10.1007/s00127-005-0882-5

Kalaycioglu, O., Copas, A., King, M., & Omar, R. Z. (2015). A comparison of multiple-imputation methods for handling missing data in repeated measurements observational studies. *Royal Statistical Society, 179*(3), 683-706.

Karahalios, A., Baglietto, L., Carlin, J. B., English, D. R., & Simpson, J. A. (2012). A review of the reporting and handling of missing data in cohort studies with repeated assessment of exposure measures. *BMC Med Res Methodol, 12*, 96.

Karahalios, A., Baglietto, L., Lee, K. J., English, D. R., Carlin, J. B., & Simpson, J. A. (2013). The impact of missing data on analyses of a time-dependent exposure in a longitudinal cohort: a simulation study. *Emerging Themes in Epidemiology, 10*(1), 6-17.

Kiess, W., Reich, A., Muller, G., Meyer, K., Galler, A., Bennek, J., & Kratzsch, J. (2001). Clinical aspects of obesity in childhood and adolescence - diagnosis, treatment and prevention. *Int J Obes (Lond), 25*(1), 75-79.

Kim, J. K., Brick, J. M., Fuller, W. A., & Kalton, G. (2006). On the bias of the multiple-imputation variance estimator in survey sampling. *Royal Statistical Society, 68*(3), 509-521.

Lee, K. J., & Carlin, J. B. (2010). Multiple Imputation for Missing Data: Fully Conditional Specification Versus Multivariate Normal Imputation. *American Journal of Epidemiology, 171*(5), 624-632.

Lumley, T. (2010). *Complex surveys : a guide to analysis using R*: Wiley.

Marshall, A., Altman, D. G., Royston, P., & Holder, R. L. (2010). Comparison of techniques for handling missing covariate data within prognostic modelling studies: a simulation study. *BMC Med Res Methodology, 10*, 7.

Mathworks. (2015). MATLAB. *Natick: Mathworks*.

This article is protected by copyright. All rights reserved.

University of Michigan. (2020). IVEware: Imputation and Variance Estimation Software, Version 0.3. Retrieved from https://www.src.isr.umich.edu/software/
Morris, T. P., White, I. R., & Crowther, M. J. (2019). Using simulation studies to evaluate statistical methods. *Statistics in Medicine, 38*(11), 2074-2102.
Nevalainen, J., Kenward, M. G., & Virtanen, S. M. (2009). Missing values in longitudinal dietary data: a multiple imputation approach based on a fully conditional specification. *Stat Med, 28*(29), 3657-3669.
Raghunathan, T. E., Lepkowsi, J. M., Van Hoewyk, J., & Solenberger, P. (2001). A multivariate technique for multiply imputing missing values using a sequence of regression models. *Survey Methodology, 27*(1), 85-95.
Reiter, J. P., Raghunathan, T. E., & Kinney, S. K. (2006). The Importance of Modeling the Sampling Design in Multiple Imputation for Missing Data. *Survey Methodology, 32*(2), 143-150.
Robins, J. M., Rotnitzky, A., & Zhao, L. P. (1994). Estimation of Regression Coefficients When Some Regressors Are Not Always Observed. *Journal of the American Statistical Association, 89*(427), 846-866.
Robins, J. M., Rotnitzky, A., & Zhao, L. P. (1995). Analysis of semiparametric regression models for repeated outcomes in the presence of missing data. *Journal of American Statistical Association, 90*(429), 106-121.
Rubin, D. B. (1987). *Multiple imputation for nonresponse in surveys*: New York ; Wiley, 1987.
Schafer, J. L. (1997). *Analysis of Incomplete Multivariate Data*. United States of America: CRC Press.
Seaman, S. R., Bartlett, J. W., & White, I. R. (2012). Multiple imputation of missing covariates with non-linear effects and interactions: an evaluation of statistical methods. *BMC Med Res Methodol, 12*, 46.
Seaman, S. R., & White, I. R. (2013). Review of inverse probability weighting for dealing with missing data. *Stat Methods Med Res, 22*(3), 278-295.
Seaman, S. R., White, I. R., Copas, A. J., & Li, L. (2012). Combining multiple imputation and inverse-probability weighting. *Biometrics, 68*(1), 129-137.
Soloff, C., D., L., & Johnstone, R. (2005). *LSAC Technical Paper No. 1: Sample design.* Australian Institute of Family Studies. (ISSN 1832-9918).
Stata Corporation. (2013). Stata statistical software, release 13. In.: College Station, TX: StataCorp LP.
Van Buuren, S., Brand, J. P. L., Groothuis-Oudshoorn, C. G. M., & Rubin, D. B. (2006). Fully conditional specification in multivariate imputation. *Journal of Statistical Computation and Simulation, 76*(12), 1049-1064.
Vansteelandt, S., Carpenter, J., & Kenward, M. G. (2010). Analysis of Incomplete Data Using Inverse Probability Weighting and Doubly Robust Estimators. *Methodology, 6*(1), 37-48.
von Hippel, P. T. (2009). How to impute interactions, squares, and other transformed variables. *Sociological Methodology, 39*, 265-291.
Welch, C., Bartlett, J., & Peterson, I. (2014). Application of multiple imputation using the two-fold fully conditional specification algorithm in longitudinal clinical data. *The Stata Journal, 14*(2), 418-431.
Welch, C., Petersen, I., Bartlett, J. W., White, I. R., Marston, L., Morris, R. W., . . . Carpenter, J. (2014). Evaluation of two-fold fully conditional specification multiple imputation for longitudinal electronic health record data. *Statistics in Medicine, 33*(21), 3725-3737.
White, I. R. (2010). simsum: Analyses of simulation studies including Monte Carlo error. *The Stata Journal, 10*(3), 369-385.
White, I. R., Royston, P., & Wood, A. M. (2011). Multiple imputation using chained equations: Issues and guidance for practice. *Statistics in Medicine, 30*(4), 377-399.
Zhou, H., Elliott, M. R., & Raghunathan, T. E. (2016). A two-step semiparametric method to accommodate sampling weights in multiple imputation. *Biometrics, 72*(1), 242-252.

Author Manuscript

This article is protected by copyright. All rights reserved.