Open Access Medical Statistics

Dovepress
open access to scientific and medical research

Open Access Full Text Article

METHODOLOGY

Comparing the performance of different multiple imputation strategies for missing binary outcomes in cluster randomized trials: a simulation study

This article was published in the following Dove Press journal:
Open Access Medical Statistics
20 December 2012
Number of times this article has been viewed

Jinhui Ma[1–3]
Parminder Raina[1,2]
Joseph Beyene[1]
Lehana Thabane[1,3–5]

[1]Department of Clinical Epidemiology and Biostatistics, McMaster University, Hamilton, ON, Canada; [2]McMaster University Evidence-based Practice Center, Hamilton, ON, Canada; [3]Biostatistics Unit, St Joseph's Healthcare Hamilton, Hamilton, ON, Canada; [4]Centre for Evaluation of Medicines, St Joseph's Healthcare Hamilton, Hamilton, ON, Canada; [5]Population Health Research Institute, Hamilton Health Sciences, Hamilton, ON, Canada

Introduction: Although researchers have proposed various strategies to handle missing outcomes in cluster randomized trials (CRTs), limited attention has been paid to the performance of these strategies. Under the assumption of covariate-dependent missingness, the objective of this simulation study is to compare the performance of various strategies in handling missing binary outcomes in CRTs under different design settings.

Methods: There are six missing data strategies investigated in this paper, which include complete case analysis, standard multiple imputation (MI) strategies using either logistic regression or Markov chain Monte Carlo (MCMC) method, within-cluster MI strategies using either logistic regression or MCMC method, and MI using logistic regression with cluster as a fixed effect. The performance of these strategies is evaluated through bias, empirical standard error, root mean squared error, and coverage probability.

Results: Under the assumption of covariate-dependent missingness and applying the generalized estimating equations approach for fitting the logistic regression, it was shown that complete case analysis yields valid inferences when the percentage of missing outcomes is not large (<20%) for all the designs of CRTs considered in this paper. Standard MI strategies can be adopted when the design effect is small (variance inflation factor [VIF] ≤ 3); however, they tend to underestimate the standard error of treatment effect when the design effect is large. Within-cluster MI strategy using logistic regression is valid for imputation of missing data from CRTs, especially when the cluster size is large (>50) and the design effect is large (VIF > 3). In contrast, within-cluster MI strategy using MCMC method may yield biased estimates of treatment effect for CRTs with small cluster size (≤50). MI using logistic regression with cluster as a fixed effect may substantially overestimate the standard error of the estimated treatment effect when the intracluster correlation coefficient is small. It may also lead to biased estimated treatment effect.

Conclusion: Findings from this simulation study provide researchers with quantitative evidence to guide selection of an appropriate strategy to deal with missing binary outcomes.

Keywords: missing data, design effect, variance inflation factor

Introduction

With the growing prominence of cluster randomized trials (CRTs) in health research, some attention has been paid to strategies for handling missing data in CRTs in the statistical community in recent years. Taljaard et al[1] evaluated imputation strategies for missing continuous outcomes in CRTs via simulation, assuming the data were missing completely at random (MCAR). They concluded that if the intracluster correlation coefficient (ICC) is small (<0.005), ignoring the clustering may yield acceptable Type I error; however, if the ICC is large, ignoring the clustering will lead to severe inflation of the Type I error.

Correspondence: Lehana Thabane
Biostatistics Unit/FSORC, 3rd Floor Martha, Room H325, St Joseph's Healthcare Hamilton, 50 Charlton Avenue East, Hamilton, ON L8N 4A6, Canada
Tel +1 905 522 1155 ext 33720
Fax +1 905 308 7212
Email thabanl@mcmaster.ca

Andridge[2] investigated the impact of fixed-effects modeling of clusters in multiple imputation (MI) for CRTs with continuous outcomes assuming outcomes are MCAR or missing at random (MAR). She showed that incorporation of clustering using fixed effects for clusters can lead to severe overestimation of variance of group means, and the overestimation is more severe when cluster sizes and ICCs are small. A previous study[3] compared several strategies for handling missing binary outcomes in CRTs under the assumption of MCAR and covariate-dependent missingness (CDM) and found that within-cluster and across-cluster MI strategies, which take into account intracluster correlation, provide more conservative treatment effect estimates compared with MI strategies which ignore the clustering effect.

Though researchers have proposed various strategies, comprehensive guidelines on the selection of the most appropriate or optimal strategy for handling missing binary outcomes from CRTs are not available in the literature. The generalizability of the conclusions from a previous study[3] to other design settings may be limited since the simulation study was based on a real dataset which has a relatively large cluster size and ICC. Moreover, different imputation strategies were compared through the odds ratios (ORs) and corresponding 95% confidence intervals (CIs) for the estimated treatment effect, and the kappa statistics for agreement between imputed datasets and the real dataset. Other evaluation criteria, such as bias, root mean squared error (RMSE), and coverage probability, are considered more informative to assess the accuracy and efficiency of different imputation strategies.

This present paper extends earlier work[3] and evaluates the performance of various strategies for missing binary outcomes in CRTs under different design settings. Under the assumption of CDM, this present paper focuses on the following strategies: complete case analysis; two standard MI strategies, ie, logistic regression and Markov chain Monte Carlo (MCMC) method; two within-cluster MI strategies, ie, logistic regression and MCMC method; and MI strategy using logistic regression with cluster as a fixed effect. Using the generalized estimating equations (GEE) approach for fitting the population-averaged model for clustered binary data, the performance of these strategies was compared using bias, RMSE, coverage probability of nominal 95% CI, and empirical standard error of the estimated treatment effect. The ultimate aim of this project is to provide researchers with quantitative evidence to guide selection of an appropriate strategy to deal with missing binary outcomes based on the design settings of CRTs and percentage of missing data.

Methods

MI has been widely applied to missing data problems. Rubin[4] described MI as a three-step process: (1) replace each missing value with a set of plausible values that represent the uncertainty about the right value to impute; (2) analyze the multiple imputed datasets using complete-data methods; and (3) combine the results from the multiple analyses, which allows uncertainty regarding the imputation to be taken into account.

This paper investigates the performance of six strategies to handle missing binary outcomes from CRTs under the assumption of CDM, ie, the probability of missing outcomes for CRTs depends only on the observed covariates. The six missing data strategies are: complete case analysis, two standard MI strategies that ignore the clustering (logistic regression and MCMC method), two within-cluster MI strategies (logistic regression and MCMC method), and MI using logistic regression with cluster as a fixed effect. All programming and analyses are implemented in SAS 9.2 (SAS Institute Inc, Cary, NC) in the simulation. The MI procedure is used to implement the MI, the GENMOD procedure is used to obtain the intervention effect estimate and its standard error from the GEE approach, and the MIANALYZE procedure is used to obtain the pooled estimate and standard error across multiple imputed datasets.

This section is organized with an introduction of the strategies investigated in this paper, followed by an illustration of the statistical method used to analyze the binary outcomes from CRTs, and finally, a description of how the results from multiple imputed datasets are combined to obtain pooled results.

Missing data strategies

Complete analysis

A complete case analysis simply omits those for whom data are incomplete. This commonly used approach loses power and may introduce bias given that the incompleteness of data is not random.

Standard MI

Logistic regression method

The standard MI using logistic regression[5] is implemented through the following steps.

1. Fit a logistic regression using the observed outcome and covariates to obtain the posterior predictive distribution of the parameters:

$$\text{logit}(\Pr(y_{obs}=1)) = \beta_0 + \beta_1 x_1 + \cdots + \beta_k x_k,$$

where y_{obs} is the observed binary outcome of a subject; x_i, $i = 1, \ldots, k$, denotes the ith individual or cluster level covariates of the corresponding subject; $\beta = (\beta_0, \beta_1, \ldots, \beta_k)$ denotes the regression coefficients; and $\text{logit}(\Pr(y_{obs} = 1)) = \log(\Pr(y_{obs} = 1)/(1 - \Pr(y_{obs} = 1)))$. In this project, only two covariates were included (treatment groups and another variable associated with the probability of missingness). The regression parameter estimates $\hat{\beta} = (\hat{\beta}_0, \hat{\beta}_1, \ldots, \hat{\beta}_k)$ and the associated covariance matrix V are obtained to construct the posterior distribution of the parameters.

2. Draw new parameters $\tilde{\beta} = (\tilde{\beta}_0, \tilde{\beta}_1, \ldots, \tilde{\beta}_k)$ from the posterior distribution, where $\tilde{\beta} = \hat{\beta} + V_h' Z$, V_h' is the upper triangular matrix in the Cholesky decomposition, $V = V_h' V_h$, and Z is a vector of $k+1$ independent random normal variates.
3. For each subject with a missing outcome y_{mis} and observed covariates $x_1, \ldots, x_k$, compute $p = \exp(\tilde{\beta}_0 + \tilde{\beta}_1 x_1 + \cdots + \tilde{\beta}_k x_k)/(1 + \exp(\tilde{\beta}_0 + \tilde{\beta}_1 x_1 + \cdots + \tilde{\beta}_k x_k))$ as the expected probability of $y_{mis} = 1$.
4. Draw a random uniform variate u, $0 \le u \le 1$. If $u < p$, then impute $y_{mis} = 1$, otherwise, impute $y_{mis} = 0$.

MCMC method

Assuming that the data are from a multivariate normal distribution, MI using MCMC method[6] constructs a Markov chain to simulate draws from the posterior distribution $\Pr(Y_{mis}|Y_{obs})$, where Y_{mis} and Y_{obs} denote the missing and observed values respectively. The missing data are imputed through repeating two steps: the imputation step and the posterior step. The ith iteration of the steps can be defined as follows.

1. Imputation step: simulate the missing values for each observation independently given an estimated mean vector and covariance matrix denoted by θ, ie, draw values for variables with missing data $Y_{mis}^{(t+1)}$ from a conditional distribution $\Pr(Y_{mis}|Y_{obs}, \theta^{(t)})$ where Y_{mis} and Y_{obs} denote variables with missing and observed data, respectively.
2. Posterior step: simulate the posterior population mean vector and covariance matrix, which are then used in the imputation step, from the complete sample estimates, ie, draw $\theta^{(t+1)}$ from $\Pr(\theta|Y_{obs}, Y_{mis}^{(t+1)})$.

The two steps are iterated long enough to generate a Markov chain $\{\theta^{(t)}, Y_{mis}^{(t)} : t = 1, 2, \cdots\}$, which converges in distribution to the posterior distribution $\Pr(Y_{mis}, \theta | Y_{obs})$.

In this study, the observed data y_{obs} include the observed outcome, treatment exposure, and the values for another variable associated with the probability of missingness. We used a single chain and noninformative prior for the Bayesian simulations to derive posterior distributions. We then applied expectation-maximization (EM) algorithm to find the maximum likelihood estimates to impute missing data. The iterations are considered to have converged when the change in the parameter estimates between iteration steps is less than 0.0001 for each parameter. Due to the assumption of multivariate normality, the imputed values from this method are continuous. They are rounded to 0 if less than 0.85, and to 1 otherwise, based on the prevalence of events in the simulated datasets.

Within-cluster MI

Within-cluster imputation refers to standard MI using either logistic regression or MCMC method being applied for each cluster independently, ie, the missing values are imputed based on the observed data within the same cluster as the missing values; therefore, the similarity of subjects from the same cluster is taken into account in within-cluster imputation methods.

Within-cluster MI strategies may not be applicable for CRTs with a small number of subjects within any cluster because the MI procedure in SAS cannot handle the case when all subjects within a cluster are missing or when the nonmissing binary outcomes within a cluster have identical observations (ie, either all 0 or all 1), phenomena that happen very often for CRTs with a small number of subjects within a cluster. In this simulation study, only the situation when all the nonmissing binary outcomes within a cluster are zero was encountered. In this case, the missing values in these clusters were replaced with zero to avoid imputing them. In addition, this strategy needs to be approached with caution, since there may be no missing data for some clusters, and in this case, the standard software programs cannot be directly applied to conduct within-cluster imputation. It will be necessary to separate clusters into two groups: clusters with missing data and clusters with complete data. For clusters with missing data, standard procedure can be used for imputing the missing data by clusters, at which point clusters with imputed data will need to be merged with the clusters with complete data for later analysis.

MI using logistic regression with cluster as a fixed effect

This method is similar to the standard MI using logistic regression; however, as its name suggests, cluster is added as a fixed effect when fitting the logistic regression using observed data and the logistic regression for imputing missing data.

Statistical analysis method

As the statistical analysis model in this study, the GEE approach was used for fitting the logistic regression, which

is a commonly used method for analyzing binary outcome in CRTs to estimate the marginal (or population-averaged) treatment effect. The GEE method, developed by Liang and Zeger,[7] can be formulated as:

$$\text{logit}(\Pr(y_{ijl}=1)) = \beta_0 + \beta_1 x_{ijl}^1 + \cdots + \beta_k x_{ijl}^k,$$

where y_{ijl} denotes the binary outcome of patient l in cluster j in the intervention group i, x_{ijl}^k denotes the kth individual-level or cluster-level covariates of the corresponding subject, β_k denotes the regression coefficients, and $\text{logit}(\Pr(y_{ijl}=1)) = \log(\Pr(y_{ijl}=1)/(1-\Pr(y_{ijl}=1)))$.

An exchangeable correlation matrix is specified to account for the potential within-cluster homogeneity in outcomes, and the robust standard error method is used to obtain the improved standard error for the estimated β coefficients. In this paper, only treatment exposure is included in the model fitting.

Another statistical analysis method, random-effects logistic regression, is also widely used to estimate the conditional (or cluster-specific) treatment effect. In this simulation study, the GEE method is adopted since the marginal treatment effect it tries to estimate is consistent with the effect used to generate the clustered binary data using beta-binomial distribution as described below. However, the GEE method underestimates the standard error of treatment effect when the number of clusters is small (<20). In this case, a small sample modification proposed by Mancl and DeRouen[8] was applied to GEE, which corrects the downward bias of the sandwich standard error estimator by multiplying it by $\sqrt{J/(J-1)}$, where J is the number of clusters in each arm.

Combining the results from different imputed data sets

For multiple imputed CRT data, the estimate of treatment effect (logarithm of the OR) and its variance are obtained in the same fashion as for the independent data. Suppose M sets of imputed values are generated, and the M estimates of the treatment effects are $\beta^{(1)}$, $\beta^{(2)}$, …, and $\beta^{(M)}$ with corresponding variance estimates $V^{(1)}, V^{(2)}$, …, and $V^{(M)}$, these estimates can be combined as described by Rubin.[5] The point estimate for the treatment effect estimate from MI is $\bar{\beta} = (\sum_{m=1}^{M} \beta^{(m)})/M$, its variance estimate is $V = W + (1+1/M)B$, where $W = (\sum_{m=1}^{M} V^{(m)})/M$ is the average within-imputation variance, and $B = (\sum_{m=1}^{M} (\beta^{(m)} - \bar{\beta})^2)/(M-1)$ is the between imputation variance. The adjusted t-test under the MI is then given by $T = (\bar{\beta} - \beta)/\sqrt{V} \sim t_{v_M}$. The degree of freedom is calculated as $v_M = (M-1)(1 + MW/(M+1)/B)^2$. For CRTs, complete data degrees of freedom are small since they are based on the number of clusters rather than the total number of subjects. In this study, the adjusted degree of freedom recommended by Barnard and Rubin[9] and Little and Rubin[10] was used, ie, $v_{adj} = \left(1/v_M + (V/W)(v_{com}+3)/(v_{com}+1)/v_{com}\right)^{-1}$, where v_{com} is the degree of freedom for the complete data test: if, for example, there are k ($k > 2$) clusters in each of the two study groups, $v_{com} = 2(k-1)$.

Simulation study

This section describes considerations for selection of design parameters for CRTs, the data generation process, and measures of performance for the missing data strategies.

Choices of design parameters for the simulation

For simplicity, only two-arm CRTs which are completely randomized, have an equal number of subjects within each cluster, and an equal number of clusters within each arm are considered. The number of clusters, the number of subjects per cluster, and the ICC are allowed to vary.

In accordance with the review of CRTs in primary care by Eldridge et al,[11] the CRTs were categorized by sample size as either trials with a small number of clusters and a large number of subjects in each cluster or trials with a large number of clusters and a small number of subjects in each cluster. The empirical findings suggest that larger values of ICC tend to be associated with studies having a small number of participants within each cluster.[12] Guided by this information, the choices of combinations of design parameters used in this simulation study are as follows.

1. CRTs with $n = 6$ clusters per arm, $m = 500$ subjects per cluster, and ICC is $\rho = 0.001$, 0.01, and 0.05.
2. CRTs with $n = 20$ clusters per arm, $m = 50$ subjects per cluster, and ICC is $\rho = 0.01$, 0.05, and 0.1.
3. CRTs with $n = 30$ clusters per arm, $m = 30$ subjects per cluster, and ICC is $\rho = 0.05$, 0.1, and 0.2.
4. The choice of the percentage of missing binary outcome is 20%; the outcome prevalence for the intervention and control arms are 10% and 20% respectively.

In addition, five replacements are generated for each of the missing data, ie, generate five datasets when applying the above MI strategies to achieve relative efficiency of more than 90%.[5]

Data generation

According to Ridout et al, clustered binomial responses can be generated using a beta-binomial distribution.[13] The prevalence of outcome π ($0 < \pi < 1$) varies from cluster to cluster according to a beta distribution with parameters $\alpha > 0$ and $\beta > 0$. The binary outcomes for each cluster are generated from the binomial distribution conditional on this prevalence or probability. To generate data with an intracluster correlation coefficient $0 < \rho < 1$ and marginal prevalence of outcome π, the parameters of the beta distribution are chosen such that $\alpha = \pi(1-\rho)/\rho$ and $\beta = (1-\pi)(1-\rho)/\rho$.

Besides the variable of intervention group, another binary covariate is generated which is associated with the probability of missingness. It is assumed that this binary covariate has equal chance to take the value of 0 or 1, and is independent of the intervention and the outcome. For any percentage of missing data, it is considered that subjects with value of 1 for this binary covariate are 1.3 times more likely to have missing outcome than subjects with value of 0. This variable is incorporated into the imputation models. Moreover, for each combination of the design parameters, 1200 replications are generated to achieve enough precision for estimating treatment effect (within 5% accuracy of the true effect for all designs of CRTs investigated in this paper with a 5% significance level).

Measures of performance

Quantities used to assess the performance of various missing data strategies include bias, RMSE, coverage probability, and standard error of the treatment effect. Details of these measures are presented below.

Bias

Bias is defined as the difference between the average value of estimated treatment effects over the simulation repetitions and the true parameter set for treatment effect when generating data.

RMSE

The mean squared error (MSE) is defined as the average squared difference between the estimated treatment effects $\hat{\beta}$ and true parameter β, which is set for treatment effect when generating the data. MSE is equal to the sum of the variance and the squared bias of the estimated treatment effect. RMSE is $\sqrt{E_\beta[(\hat{\beta}-\beta)^2]}$, which is the square root of the MSE. The RMSE is a useful measure of overall precision or accuracy.

Coverage

The actual coverage of nominal 95% CIs of the estimated treatment effect is the proportion of time that the nominal interval contains the true treatment effect across all simulation replications. Since the 95% CI aims to contain the true treatment effect with probability of 0.95, nominal coverage should be approximately equal to coverage probability if the missing data strategy works well.

Empirical standard error of the treatment effect

Empirical standard error of the treatment effect is calculated as the average of standard errors of estimated treatment effects across all simulation replications. It has been well established that analytical methods failing to account for the correlation between responses within cluster, ie, the clustering effect, result in underestimation of standard error for the intervention effect. The appropriate imputation model should also reflect this data structure, as pointed out by Kenward and Carpenter;[14] therefore, the empirical standard error is considered to be the primary criterion in this study.

Results

Since subjects within the same cluster are more likely to be similar to each other than those from different clusters, an additional subject from the same cluster adds less new information than would a completely independent subject. The design effect or the variance inflation factor (VIF) is commonly used to measure the clustering effect due to lack of independence in the data from a CRT design. The main components of the design effect are the intracluster correlation coefficient ρ and the size of cluster m, and VIF = $1 + (m-1)\rho$. For each design setting of CRT investigated in this simulation study, the empirical standard error of estimated treatment effect, bias, RMSE, and coverage probability for analyzing the complete data (no missing data) are considered as references and compared with those for each missing data strategy. These results are presented in Tables 1–6 and discussed in detail below.

Complete case analysis

For all the design settings of CRTs, the empirical standard errors are inflated slightly; the biases are close to zero; and the RMSEs and coverage probabilities are very similar to their references (see Table 1). This result is not surprising since the complete case analysis is analogous to analyzing a size-reduced dataset in which all variables are fully observed under the assumption of CDM. It can yield an unbiased estimate of the intervention

Table 1 Performance of complete case analysis

Design parameters of CRTs			Design effect (variance inflation factor)	Empirical standard error		Bias		RMSE		Coverage	
Number of clusters per arm (m)	Number of subjects per cluster (n)	Intracluster correlation coefficient (ρ)		No missing data (Ref)	Complete case analysis	No missing data (Ref)	Complete case analysis	No missing data (Ref)	Complete case analysis	No missing data (Ref)	Complete case analysis
6[a]	500	0.001	1.499	0.08/0.09	0.09/0.10	0.00	0.00	0.10	0.10	0.88/0.91	0.89/0.92
		0.01	5.99	0.17/0.18	0.17/0.19	−0.01	−0.01	0.18	0.18	0.92/0.94	0.91/0.94
		0.05	25.95	0.34/0.38	0.34/0.38	−0.03	−0.03	0.38	0.38	0.91/0.94	0.90/0.94
20	50	0.01	1.49	0.16	0.17	0.01	0.01	0.16	0.18	0.92	0.93
		0.05	3.45	0.24	0.25	−0.02	−0.02	0.24	0.24	0.95	0.96
		0.1	5.9	0.31	0.32	−0.04	−0.04	0.32	0.33	0.94	0.94
30	30	0.05	2.45	0.21	0.22	−0.01	−0.02	0.22	0.23	0.94	0.94
		0.1	3.9	0.27	0.28	−0.00	−0.01	0.27	0.28	0.94	0.94
		0.2	6.8	0.35	0.36	−0.03	−0.03	0.37	0.38	0.93	0.93

Note: [a]For CRTs with six clusters per arm, standard errors/modified standard errors and coverage/modified coverage are provided.
Abbreviations: CRT, cluster randomized trial; Ref, reference; RMSE, root mean squared error.

Table 2 Performance of standard MI using LR

Design parameters of CRTs			Design effect (variance inflation factor)	Empirical standard error		Bias		RMSE		Coverage	
Number of clusters per arm (m)	Number of subjects per cluster (n)	Intracluster correlation coefficient (ρ)		No missing data (Ref)	Standard MI using LR	No missing data (Ref)	Standard MI using LR	No missing data (Ref)	Standard MI using LR	No missing data (Ref)	Standard MI using LR
6[a]	500	0.001	1.499	0.08/0.09	0.09/0.10	0.00	0.00	0.10	0.10	0.88/0.91	0.89/0.93
		0.01	5.99	0.17/0.18	0.14/0.16	−0.01	−0.01	0.18	0.18	0.92/0.94	0.86/0.91
		0.05	25.95	0.34/0.38	0.28/0.31	−0.03	−0.03	0.38	0.38	0.91/0.94	0.83/0.87
20	50	0.01	1.49	0.16	0.16	0.01	0.01	0.16	0.18	0.92	0.92
		0.05	3.45	0.24	0.22	−0.02	−0.02	0.24	0.25	0.95	0.92
		0.1	5.9	0.31	0.27	−0.04	−0.04	0.32	0.33	0.94	0.89
30	30	0.05	2.45	0.21	0.20	−0.01	−0.02	0.22	0.23	0.94	0.91
		0.1	3.9	0.27	0.24	−0.00	−0.01	0.27	0.28	0.94	0.90
		0.2	6.8	0.35	0.31	−0.03	−0.03	0.37	0.38	0.93	0.89

Note: [a]For CRTs with six clusters per arm, standard errors/modified standard errors and coverage/modified coverage are provided.
Abbreviations: CRT, cluster randomized trial; LR, logistic regression; MI, multiple imputation; Ref, reference; RMSE, root mean squared error.

Table 3 Performance of standard MI using MCMC method

Design parameters of CRTs			Design effect (variance inflation factor)	Empirical standard error		Bias		RMSE		Coverage	
Number of clusters per arm (m)	Number of subjects per cluster (n)	Intracluster correlation coefficient (ρ)		No missing data (Ref)	Standard MI using MCMC	No missing data (Ref)	Standard MI using MCMC	No missing data (Ref)	Standard MI using MCMC	No missing data (Ref)	Standard MI using MCMC
6[a]	500	0.001	1.499	0.08/0.09	0.09/0.10	0.00	0.02	0.10	0.10	0.88/0.91	0.88/0.91
		0.01	5.99	0.17/0.18	0.16/0.18	−0.01	0.01	0.18	0.18	0.92/0.94	0.90/0.93
		0.05	25.95	0.34/0.38	0.32/0.36	−0.03	−0.01	0.38	0.37	0.91/0.94	0.89/0.93
20	50	0.01	1.49	0.16	0.16	0.01	0.03	0.16	0.18	0.92	0.93
		0.05	3.45	0.24	0.24	−0.02	0.00	0.24	0.24	0.95	0.95
		0.1	5.9	0.31	0.30	−0.04	−0.01	0.32	0.32	0.94	0.93
30	30	0.05	2.45	0.21	0.22	−0.01	0.00	0.22	0.22	0.94	0.94
		0.1	3.9	0.27	0.27	−0.00	0.02	0.27	0.28	0.94	0.93
		0.2	6.8	0.35	0.34	−0.03	−0.00	0.37	0.37	0.93	0.92

Note: [a]For CRTs with six clusters per arm, standard errors/modified standard errors and coverage/modified coverage are provided.
Abbreviations: CRT, cluster randomized trial; MCMC, Markov chain Monte Carlo; MI, multiple imputation; Ref, reference; RMSE, root mean squared error.

Table 4 Performance of within-cluster MI using LR

Design parameters of CRTs			Design effect (variance inflation factor)	Empirical standard error		Bias		RMSE		Coverage	
Number of clusters per arm (m)	Number of subjects per cluster (n)	Intracluster correlation coefficient (ρ)		No missing data (Ref)	Within-cluster MI using LR	No missing data (Ref)	Within-cluster MI using LR	No missing data (Ref)	Within-cluster MI using LR	No missing data (Ref)	Within-cluster MI using LR
6[a]	500	0.001	1.499	0.08/0.09	0.10/0.12	0.00	0.00	0.10	0.11	0.88/0.91	0.94/0.96
		0.01	5.99	0.17/0.18	0.18/0.20	−0.01	−0.01	0.18	0.18	0.92/0.94	0.92/0.95
		0.05	25.95	0.34/0.38	0.35/0.39	−0.03	−0.03	0.38	0.38	0.91/0.94	0.90/0.94
20	50	0.01	1.49	0.16	0.19	0.01	0.05	0.16	0.18	0.92	0.96
		0.05	3.45	0.24	0.26	−0.02	0.02	0.24	0.24	0.95	0.97
		0.1	5.9	0.31	0.33	−0.04	−0.01	0.32	0.32	0.94	0.95
30	30	0.05	2.45	0.21	0.24	−0.01	0.02	0.22	0.22	0.94	0.97
		0.1	3.9	0.27	0.29	−0.00	0.02	0.27	0.27	0.94	0.96
		0.2	6.8	0.35	0.37	−0.03	−0.01	0.37	0.37	0.93	0.94

Note: [a]For CRTs with six clusters per arm, standard errors/modified standard errors and coverage/modified coverage are provided.
Abbreviations: CRT, cluster randomized trial; LR, logistic regression; MI, multiple imputation; Ref, reference; RMSE, root mean squared error.

Table 5 Performance of within-cluster MI using MCMC method

Design parameters of CRTs			Design effect (variance inflation factor)	Empirical standard error		Bias		RMSE		Coverage	
Number of clusters per arm (m)	Number of subjects per cluster (n)	Intracluster correlation coefficient (ρ)		No missing data (Ref)	Within-cluster MI using MCMC	No missing data (Ref)	Within-cluster MI using MCMC	No missing data (Ref)	Within-cluster MI using MCMC	No missing data (Ref)	Within-cluster MI using MCMC
6[a]	500	0.001	1.499	0.08/0.09	0.10/0.11	0.00	−0.03	0.10	0.11	0.88/0.91	0.88/0.92
		0.01	5.99	0.17/0.18	0.18/0.20	−0.01	−0.04	0.18	0.20	0.92/0.94	0.91/0.94
		0.05	25.95	0.34/0.38	0.35/0.39	−0.03	−0.05	0.38	0.40	0.91/0.94	0.90/0.93
20	50	0.01	1.49	0.16	0.18	0.01	0.02	0.16	0.18	0.92	0.94
		0.05	3.45	0.24	0.24	−0.02	0.03	0.24	0.25	0.95	0.93
		0.1	5.9	0.31	0.30	−0.04	0.16	0.32	0.35	0.94	0.90
30	30	0.05	2.45	0.21	0.21	−0.01	0.17	0.22	0.27	0.94	0.89
		0.1	3.9	0.27	0.26	−0.00	0.27	0.27	0.37	0.94	0.82
		0.2	6.8	0.35	0.33	−0.03	0.37	0.37	0.51	0.93	0.77

Note: [a]For CRTs with six clusters per arm, standard errors/modified standard errors and coverage/modified coverage are provided.
Abbreviations: CRT, cluster randomized trial; MCMC, Markov chain Monte Carlo; MI, multiple imputation; Ref, reference; RMSE, root mean squared error.

Table 6 Performance of MI using LR with cluster as a fixed effect

Design parameters of CRTs			Design effect (variance inflation factor)	Empirical standard error		Bias		RMSE		Coverage	
Number of clusters per arm (m)	Number of subjects per cluster (n)	Intracluster correlation coefficient (ρ)		No missing data (Ref)	MI using LR with cluster as a fixed effect	No missing data (Ref)	MI using LR with cluster as a fixed effect	No missing data (Ref)	MI using LR with cluster as a fixed effect	No missing data (Ref)	MI using LR with cluster as a fixed effect
6[a]	500	0.001	1.499	0.08/0.09	0.33/0.37	0.00	0.06	0.10	0.15	0.88/0.91	1.00/1.00
		0.01	5.99	0.17/0.18	0.36/0.40	−0.01	0.05	0.18	0.21	0.92/0.94	0.99/1.00
		0.05	25.95	0.34/0.38	0.45/0.51	−0.03	0.03	0.38	0.38	0.91/0.94	0.97/0.98
20	50	0.01	1.49	0.16	0.21	0.01	0.05	0.16	0.18	0.92	0.96
		0.05	3.45	0.24	0.27	−0.02	0.08	0.24	0.23	0.95	0.97
		0.1	5.9	0.31	0.32	−0.04	0.12	0.32	0.30	0.94	0.95
30	30	0.05	2.45	0.21	0.24	−0.01	0.15	0.22	0.24	0.94	0.94
		0.1	3.9	0.27	0.28	−0.00	0.21	0.27	0.31	0.94	0.92
		0.2	6.8	0.35	0.33	−0.03	0.25	0.37	0.38	0.93	0.91

Note: [a]For CRTs with six clusters per arm, standard errors/modified standard errors and coverage/modified coverage are provided.
Abbreviations: CRT, cluster randomized trial; LR, logistic regression; MI, multiple imputation; Ref, reference; RMSE, root mean squared error.

effect, but with a larger empirical standard error for CRTs with small design effect compared to those with large design effect; this is due to loss of efficiency.

These findings suggest that, for covariate-dependent missingness, complete case analysis may be an acceptable strategy as long as the percentage of missing data is not large (<20%). The advantage of complete case analysis lies in its simplicity, and it is the default method applied to handling missing data in the standard software.

Standard MI strategies

When standard MI using logistic regression is used to handle the missing data, the empirical standard errors are substantially underestimated for CRTs with large design effect (VIF > 3) and similar for CRTs with small design effect; the biases are close to zero; biases and RMSEs are similar to their references; and the coverage for CRTs with large design effects is smaller than their references (see Table 2). The performance of the standard MI using MCMC method is similar to that of the standard MI using logistic regression, except that the underestimation of the standard error for imputation using MCMC method is not substantial (see Table 3).

Two reasons can help to interpret why the biases for standard MI using logistic regression are close to zero: first, CDM is assumed in this simulation study; and second, both the imputation strategy and statistical analysis model (GEE approach for fitting logistic regression) estimate the population-averaged treatment effect, which is consistent with the treatment effect used to generate the clustering data.

For CRTs with very small design effect, the information contributed from a subject within a cluster is quite similar to that from a completely independent subject and therefore the standard MI using logistic regression, which accounts for the uncertainty of the missing data through both within- and between-imputation variances, may provide larger standard error compared with that when there are no missing data. However, for CRTs with large design effects (VIF > 3), this strategy underestimates standard error of the intervention effect since it ignores the clustering of data.

These findings suggest that the standard MI strategies are acceptable when the VIF is small (<3); otherwise, they tend to underestimate the standard error of the treatment effect. In addition, an MI strategy using MCMC can be applied for arbitrary missing data pattern, whereas an MI strategy using logistic regression can only be applied for monotone missing pattern. MI strategy using the MCMC method presents a convergence problem and may produce biased results when the prevalence of cases (ie, the prevalence of outcome) is close to 0 or to 1.

Within-cluster MI strategies

The empirical standard errors for within-cluster MI using logistic regression are larger than their references, and the increased amount is ignorable for CRTs with large design effect and large cluster size; the biases and the RMSEs are quite similar to their references; the coverage probabilities are slightly larger than their references (see Table 4). This strategy imputes each incident of missing data based on only the observed data within the same cluster, which leads to increasing within- and between-imputation variance and thus affects overall variance of estimated treatment effect when compared with imputation based on all observed data across the different clusters.

The empirical standard errors for within-cluster MI using MCMC method are similar to their references; for CRTs with small cluster size, the biases are not ignorable, which lead to larger RMSEs and smaller coverage probabilities compared with their references (see Table 5).

These findings suggest that within-cluster MI strategies are an appropriate imputation strategy for CRTs, especially for CRTs with large cluster size and large design effect (VIF > 3).

MI using logistic regression with cluster as a fixed effect

When the MI using logistic regression with cluster as a fixed effect is applied to impute the missing data, the standard errors are substantially overestimated for CRTs with small ICC($P < 0.1$), which lead to larger coverage probabilities compared with their references. The biases are large, especially for CRTs with small cluster size (≤50), which lead to smaller coverage probabilities. The large biases and overestimated standard errors lead to increased RMSEs compared with their references (see Table 6).

These findings suggest that application of this strategy may result in a biased estimate of treatment effect and may substantially overestimate the standard error of the estimated treatment effect when ICC is small ($\rho < 0.1$) and the cluster size is small (≤50).

Discussion

Missing data is a common issue in CRTs, which may lead to spurious conclusions if handled inappropriately. This study used an extensive set of simulations to assess the performance of different strategies for handling missing

binary outcomes in CRTs under different design settings. Results from the present study demonstrate that the design of CRTs, including factors such as the number of clusters in each intervention group, the number of subjects within each cluster, the ICC, and the VIF, are important determinants for selecting an appropriate missing data strategy. Under the assumption of CDM and application of the GEE approach for statistical analysis, complete case analysis can be used to obtain valid inference when the percentage of missing binary outcomes is small (<20%). Standard MI using logistic regression or MCMC method can be used to impute the missing values when the design effect is small (VIF $\leq$ 3); however, they tend to underestimate the standard error of the treatment effect when the design effect is large, though the underestimation of the standard MI using MCMC method is not substantial. Within-cluster MI using logistic regression may be an appropriate strategy to impute missing binary outcomes in CRTs, especially for CRTs with large cluster size and design effect. The performance of within-cluster MI using MCMC method is good for CRTs with large cluster size and design effect (VIF > 3); however, may yield biased estimation of the treatment effect for CRTs with small cluster size. The MI using logistic regression with cluster as a fixed effect substantially overestimates the standard error of the treatment effect for CRTs with small ICC (<0.05) and may result in a biased estimated treatment effect for CRTs with small cluster size.

The finding for the MI using logistic regression with cluster as a fixed effect in this paper parallels previous results by Andridge[2] who demonstrated that incorporating clusters as fixed effects to handle missing continuous outcomes can lead to serious overestimation of variance of group means, and this overestimation is more severe for small cluster sizes and small ICCs. The findings for complete case analysis, standard MI strategies which ignore the clustering effect, and within-cluster MI strategies are similar to those from Taljaard et al;[1] although, they evaluated imputation strategies for missing continuous outcomes in CRTs assuming the missingness was completely at random and used Type I error rate and statistical power as the main evaluation criteria. This present study adopted the design effect VIF and the ICC, rather than the ICC alone, to interpret simulation results, since VIF is determined by both the number of subjects within each cluster and the magnitude of ICC, and is more appropriate for capturing the pattern of the performance for different missing data imputation strategies.

It should be emphasized that complete case analysis may not be an appropriate strategy in practice though it shows good performance in this simulation study. In fact, the good performance of complete case analysis is highly dependent on the CDM assumption. In realistic scenarios, it is more likely for a CRT to have mixed missing data mechanisms, ie, combination of missing completely at random (a participant accidentally missed the medical appointment for assessing his health outcome), CDM (older participants are more likely to be lost to follow-up), or missing not at random (participants with poor health outcome are more likely to be lost to follow-up). A simulation study by Allison[15] shows that MI is robust to model violations while complete case analysis is not. King et al[16] further shows that MI works well even when the assumptions of MI are violated.

There are certain limitations to the current study. First, the performance of different missing data strategies is only assessed in the setting of a completely randomized study design. Other designs such as matched pairs design and stratified randomized design are also used for CRTs but were not considered in this study. Second, only balanced design of CRTs (with equal number of subjects per cluster, equal number of clusters in each arm) was considered. These design restrictions were made in order to understand the performance of the methods in simple scenarios. However, the findings are relevant to more general settings, such as the unbalanced design of CRTs. Third, the scenario of missing an entire cluster was not investigated. Even though the complete case analysis, standard MI strategies, and MI using logistic regression with cluster as fixed effect can manage conditions when an entire cluster is missing, their performance in this scenario needs further investigation. Finally, only the GEE approach is considered as the analysis method for the present study; however, in practice, random-effects logistic regression is also commonly adopted for analyzing binary outcomes in CRTs. Further study could include investigation of the performance of these missing data strategies when random-effects logistic regression is used as the analysis model.

The strengths of this study include comparison to a previous simulation study[3] which focused on the estimate of the treatment effect under one particular design setting while emphasis of the present study has been on the accuracy and effectiveness of different missing data strategies, and should provide more informative criteria to assess performance of different imputation strategies. As well, this simulation study is designed to cover a wide range of design settings for CRTs and applies an amount of missing data commonly encountered in epidemiological research. All the above strengths enhance the generalizability of these findings.

Conclusion

The current study is the most comprehensive to date to examine performance of different strategies for handling missing binary outcomes in CRTs. When the percentage of missing data is large and the design effect of the CRT varies, different strategies may lead to varying results, and therefore the appropriate strategy needs to be chosen carefully to obtain valid inferences and mitigate design issues. Findings from this simulation study provide researchers with quantitative evidence to guide selection of appropriate strategies for handling missing binary outcomes based on the design settings of CRTs and the percentage of missing data.

Acknowledgments

This study was supported in part by funds from the Canadian Network and Centre for Trials Internationally (CANNeCTIN) program and the Drug Safety and Effectiveness Cross-disciplinary Training (DSECT) program in the form of studentship on training awards for the first author.

Disclosure

No additional external funding was received for this study. Dr Lehana Thabane is a clinical trials mentor for the Canadian Institutes of Health Research (CIHR). Dr Parminder Raina holds a Raymond and Margaret Labarge Chair in Research and Knowledge Application for Optimal Aging. Dr Parminder Raina is the Canada research chair in GeroScience, CIHR. The authors report no conflicts of interest in this work.

References

1. Taljaard M, Donner A, Klar N. Imputation strategies for missing continuous outcomes in cluster randomized trials. *Biom J*. 2008;50(3):329–345.
2. Andridge RR. Quantifying the impact of fixed effects modeling of clusters in multiple imputation for cluster randomized trials. *Biom J*. 2011;53(1):57–74.
3. Ma J, Akhtar-Danesh N, Dolovich L, Thabane L; CHAT investigators. Imputation strategies for missing binary outcomes in cluster randomized trials. *BMC Med Res Methodol*. 2011;11:18.
4. Rubin DB. Multiple imputation after 18+ years. *J Am Stat Assoc*. 1996;91:473–489.
5. Rubin DB. *Multiple Imputation for Nonresponse in Surveys*. New York: John Wiley & Sons; 1987.
6. Schafer JL. *Analysis of Incomplete Multivariate Data*. London: Chapman & Hall; 1997.
7. Liang K, Zeger S. Longitudinal data analysis using generalized linear models. *Biometrika*. 1986;73(1):13–22.
8. Mancl LA, DeRouen TA. A covariance estimator for GEE with improved small-sample properties. *Biometrics*. 2001;57:126–134.
9. Barnard J, Rubin DB. Small-sample degrees of freedom with multiple imputation. *Biometrika*. 1999;86:949–955.
10. Little RJ, Rubin DB. Statistical analysis with missing data. 2nd ed. New York: John Wiley & Sons; 2002.
11. Eldridge SM, Ashby D, Feder GS, Rudnicka AR, Ukoumunne OC. Lessons for cluster randomized trials in the twenty-first century: a systematic review of trials in primary care. *Clin Trials*. 2004;1(1):80–90.
12. Donner A. An empirical study of cluster randomization. *Int J Epidemiol*. 1982;11(3):283–286.
13. Ridout MS, Demetrio CG, Firth D. Estimating intraclass correlation for binary data. *Biometrics*. 1999;55(1):137–148.
14. Kenward MG, Carpenter J. Multiple imputation: current perspectives. *Stat Methods Med Res*. 2007;16(3):199–218.
15. Allison P. Multiple imputation for missing data: a cautionary tale. *Sociol Methods Res*. 2000;28(3):301–309.
16. King G, Honaker J, Joseph A, Scheve K. Analyzing incomplete political science data: an alternative algorithm for multiple imputation. *Am Polit Sci Rev*. 2001;95(1):49–69.