

# Handling of outcome missing data dependent on measured or unmeasured background factors in micro-randomized trial: Simulation and application study

DIGITAL HEALTH  
Volume 10: 1–15  
© The Author(s) 2024  
Article reuse guidelines:  
sagepub.com/journals-permissions  
DOI: 10.1177/20552076241249631  
journals.sagepub.com/home/dhj



Masahiro Kondo<sup>1,2</sup>  and Koji Oba<sup>3,4</sup>

## Abstract

**Background:** Micro-randomized trials (MRTs) enhance the effects of mHealth by determining the optimal components, timings, and frequency of interventions. Appropriate handling of missing values is crucial in clinical research; however, it remains insufficiently explored in the context of MRTs. Our study aimed to investigate appropriate methods for missing data in simple MRTs with uniform intervention randomization and no time-dependent covariates. We focused on outcome missing data depending on the participants' background factors.

**Methods:** We evaluated the performance of the available data analysis (AD) and the multiple imputation in generalized estimating equations (GEE) and random effects model (RE) through simulations. The scenarios were examined based on the presence of unmeasured background factors and the presence of interaction effects. We conducted the regression and propensity score methods as multiple imputation. These missing data handling methods were also applied to actual MRT data.

**Results:** Without the interaction effect, AD was biased for GEE, but there was almost no bias for RE. With the interaction effect, estimates were biased for both. For multiple imputation, regression methods estimated without bias when the imputation models were correct, but bias occurred when the models were incorrect. However, this bias was reduced by including the random effects in the imputation model. In the propensity score method, bias occurred even when the missing probability model was correct.

**Conclusions:** Without the interaction effect, AD of RE was preferable. When employing GEE or anticipating interactions, we recommend the multiple imputation, especially with regression methods, including individual-level random effects.

## Keywords

Micro-randomized trial, missing data, multiple imputation, mobile health, mobile app

Submission date: 18 December 2023; Acceptance date: 8 April 2024

## Introduction

The widespread adoption of smartphones and other mobile devices has driven considerable attention towards mobile health for diagnosis, treatment, and prevention of diseases. As of 2018, over 300,000 health-related applications have been developed,<sup>1,2</sup> typically based on the best possible behavior change theory. However, with few exceptions, it remains challenging to determine the optimal timings and frequency of interventions based on these theories.<sup>3,4</sup> Micro-randomized trials (MRTs) have been introduced as a trial design to optimize apps by obtaining required

<sup>1</sup>Biostatistics Unit, Clinical and Translational Research Center, Keio University Hospital, Tokyo, Japan

<sup>2</sup>Graduate School of Health Management, Keio University, Kanagawa, Japan

<sup>3</sup>Interfaculty Initiative in Information Studies, the University of Tokyo, Tokyo, Japan

<sup>4</sup>Department of Biostatistics, School of Public Health, Graduate School of Medicine, the University of Tokyo, Tokyo, Japan

### Corresponding author:

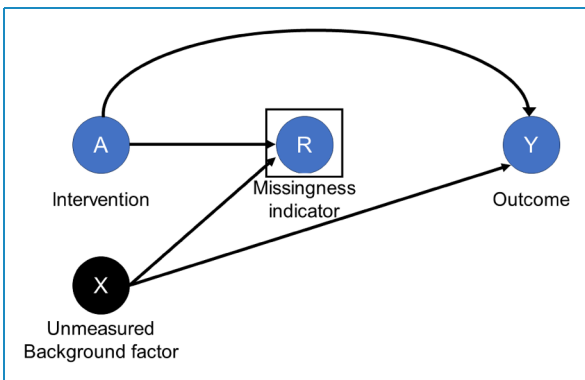
Koji Oba, Department of Biostatistics, School of Public Health, Graduate School of Medicine, the University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo, Japan.

Email: oba@epistat.m.u-tokyo.ac.jp



knowledge through clinical studies.<sup>4</sup> MRTs usually span several weeks, up to a few months, with participants randomized numerous times at pre-determined time points. In MRTs, the assignment of the interventions could be changed depending on previous outcomes and time-dependent covariates. Time-dependent covariates would also be considered when evaluating the intervention effects.<sup>4,5</sup> Although study design incorporating time-dependent covariates and previous outcomes is the strength of MRT, some trials have been more simply designed and analyzed.<sup>6–10</sup> Trials where time-dependent covariates or previous outcomes are used only to increase statistical power in the analysis<sup>11</sup> can be considered simple MRTs.

Missing data are a frequent issue in clinical studies of health-related apps. The systematic review of apps for chronic diseases reported that approximately 40% of data were missing, on average.<sup>12</sup> The missing data are often classified as missing completely at random (MCAR), missing at random (MAR), and missing not at random (MNAR), with these classifications varying across.<sup>13–17</sup> For our purposes, we adopted the definitions used in the paper by Curnow et al.<sup>17</sup> We simplified their definition and applied them to our study. MAR was defined as the case in which missingness depends on only observed variables; MNAR was defined as the case in which missingness depends on unobserved variables. Although research on the causes of missing data in clinical studies of apps is limited, the associated factors are considered to be diverse, including age, educational history, occupation, health literacy, and personality characteristics.<sup>12</sup> These factors are numerous, and some are difficult to measure; thus, measuring all factors associated with missing data is typically challenging. In addition, variables such as personality characteristics are associated with lifestyles, like the number of walking steps.<sup>18–20</sup> Therefore, unmeasured background factors may introduce non-negligible bias into the estimates (Figure 1).



**Figure 1.** Directed acyclic graph illustrating non-negligible bias generation mechanisms due to unmeasured background factors.

To the best of our knowledge, appropriate methods for handling missing data in MRTs have not yet been adequately evaluated. We focused on a simple MRT in which interventions were randomized in the same probability for all participants and without considering time-dependent covariates in the analysis. As a first step in considering missing data in MRT, our study aimed to investigate the proper methods for handling missing data in MRTs when missingness depends on background factors. We performed two generally preferred methods, the available data analysis and the multiple imputation.<sup>16</sup> Sixteen scenarios were considered, based on presence of the interaction effect between the background factor and the intervention, presence of unmeasured background factors, types of background factors (continuous or categorical), and influence of covariates on probabilities of the missing (large or small). Following the simulation study, we applied missing data handling methods to the MRT of the assistant to lift your level of activity (Ally), an exercise promotion app for healthy adults.<sup>8</sup>

## Methods

In repeated measures studies, generalized estimating equations (GEE) and the random effects model (RE) are often used for estimating the intervention effects.<sup>21</sup> GEE and RE were not valid for MRTs with time-dependent covariates,<sup>5,22</sup> but we used GEE and RE in our analysis since we assumed a situation in which such time-dependent variables were not present. For the handling of missing measurements, we performed the available data analysis and the multiple imputation.

### Generalized estimating equations<sup>21</sup>

GEE is a multivariate analysis method for clustered data, for instance, where subjects are measured at many time points. We estimate the intervention's or other variables' effect ( $\beta$ ) by solving the following estimating equation:

$$\sum_{i=1}^n \left( \frac{\partial \mu_i(\beta)}{\partial \beta} \right)^t V_i^{-1} (Y_i - \mu_i(\beta)) = 0$$

where  $Y_i$  is the vector of outcomes for subject  $i$ ,  $E(Y_i) = \mu_i(\beta) = g(X_i\beta)$ ,  $X_i$  is the design matrix including the intervention and other variables,  $g(\cdot)$  is the link-function such as identify, and  $V_i = \text{Var}(Y_i)$ . When using GEE, it is necessary to specify the  $\mu_i(\beta)$  model and the  $V_i$  structure. The following sandwich variances provide estimators that are robust to misspecification of the structure of  $V_i$ .

$$\text{Var}(\hat{\beta}) = D^{-1} M D^{-1}$$

$$D = \sum_{i=1}^n \left( \frac{\partial \mu_i(\beta)}{\partial \beta} \right)^t V_i^{-1} (Y_i - \mu_i(\beta))$$

$$\mathbf{M} = \sum_{i=1}^n \left( \frac{\partial \boldsymbol{\mu}_i(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}} \right)^t \mathbf{V}_i^{-1} (\mathbf{Y}_i - \boldsymbol{\mu}_i(\boldsymbol{\beta}))^t (\mathbf{Y}_i - \boldsymbol{\mu}_i(\boldsymbol{\beta})) \mathbf{V}_i^{-1} \left( \frac{\partial \boldsymbol{\mu}_i(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}} \right)$$

### Random effects model<sup>21</sup>

RE is also a multivariate analysis method for clustered data. However, unlike GEE, the effect is estimated by maximum likelihood estimation. We describe the linear model that we performed in this study. Outcomes are modeled as follows:

$$Y_{it} = \mathbf{X}_{it}^t \boldsymbol{\beta} + \mathbf{Z}_{it}^t \mathbf{b}_i + \epsilon_{it}$$

where  $Y_{it}$  is the outcome for subject  $i$  at time-point  $t$ ,  $\mathbf{X}$  is the fixed effects covariates,  $\mathbf{Z}$  is the random effects covariates,  $\mathbf{b}_i$  is the random effect, and  $\epsilon_{it}$  is the error term. We assume that  $\epsilon_{it}$  follows an independent normal distribution and  $\mathbf{b}_i$  follows a multivariate normal distribution.

### Available data analysis

A simple approach to address missing data in these analyses is to remove missing data and analyze only the observed data. This method is known as the available data analysis. In general, non-negligible bias occurs in the available data analysis when unmeasured covariates affect missingness and outcomes.<sup>23</sup> In this study, we performed the available data analysis on GEE or RE and compared the results with the case of the multiple imputation.

### Multiple imputation

The multiple imputation is a statistical approach for addressing missing data by repeatedly imputing missing values. The multiple imputation consists of three steps. The first step is to create  $m$  sets of pseudo-complete datasets by imputing repeatedly missing values based on the imputation models. The second step is to estimate the parameters  $\boldsymbol{\theta}$  and their covariance matrixes by analyzing each  $m$  set of the pseudo-complete datasets generated in the first step. The estimated values of  $\boldsymbol{\theta}$  are denoted as  $\hat{\boldsymbol{\theta}}_1, \hat{\boldsymbol{\theta}}_2, \dots, \hat{\boldsymbol{\theta}}_m$  and the estimated covariance matrixes as  $\hat{\mathbf{V}}_1, \hat{\mathbf{V}}_2, \dots, \hat{\mathbf{V}}_m$ . The final step is to combine the estimates obtained in the second step into one: the estimator of  $\boldsymbol{\theta}$  is the mean of estimates, and the estimator of the covariance matrix is given by the Rubin rule as follows.  $\mathbf{W}_{IM}$  is the within-variance, and  $\mathbf{B}_{IM}$  is the between-variance.

$$\bar{\boldsymbol{\theta}} = \frac{1}{m} \sum_{l=1}^m \hat{\boldsymbol{\theta}}_l$$

$$V(\bar{\boldsymbol{\theta}}) = \mathbf{W}_{IM} + \left(1 + \frac{1}{m}\right) \mathbf{B}_{IM}$$

$$\mathbf{W}_{IM} = \frac{1}{m} \sum_{l=1}^m \hat{\mathbf{V}}_l, \mathbf{B}_{IM} = \frac{1}{m-1} \sum_{l=1}^m (\hat{\boldsymbol{\theta}}_l - \bar{\boldsymbol{\theta}})(\hat{\boldsymbol{\theta}}_l - \bar{\boldsymbol{\theta}})^T$$

The multiple imputation was initially proposed based on the Markov chain Monte Carlo method, which randomly imputes missing values from the posterior predictive distribution conditioned on the observed values.<sup>24</sup> However, implementing random sampling is difficult. Thus, a more flexible approach based on fully conditional specification (FCS) was proposed.<sup>25</sup> In this approach, missing data imputation is based on variable-by-variable models under the assumption that the conditional distribution of each variable is defined by the other variables. Let  $\mathbf{X}$  consist of  $p$  variables measured and  $\mathbf{R}$  be the missing indicator for  $\mathbf{X}$ , and they are drawn from the multivariate distribution  $P(\mathbf{X}, \mathbf{R} | \boldsymbol{\varphi})$ . We assume all  $X_j$ ,  $j = 1, 2, 3, \dots, p$  could be specified by  $\mathbf{X}_{-j}$ , which is  $\mathbf{X}$  without  $X_j$ . FCS specifies a set of conditional distributions  $P(X_j | \mathbf{X}_{-j}, \mathbf{R}, \boldsymbol{\varphi}_j)$  instead of  $P(\mathbf{X}, \mathbf{R} | \boldsymbol{\varphi})$ . We evaluated the performance of the multiple imputation based on the FCS in MRTs. Specifically, the regression method<sup>26</sup> and the propensity score method<sup>27</sup> were performed.

**Regression method<sup>26</sup>.** In the regression method, a regression analysis with the imputed variable as the explained variable is performed using the observed data. The model

$$X_j = \mathbf{X}_{-j}^t \boldsymbol{\beta}_{\text{REG}}$$

is fitted. The posterior predictive distributions of the regression parameters are calculated from the obtained estimates. The new parameter  $\boldsymbol{\beta}_*$  obtained based on  $\hat{\boldsymbol{\beta}}_{\text{REG}}$  and its variance structure. The posterior predictive distribution of the missing values  $X_j$  is obtained based on the regression parameters randomly drawn from the posterior predictive distribution. The missing values are randomly imputed from the distribution with an expected value of  $\mathbf{X}_{-j}^t \boldsymbol{\beta}_*$ .

**Propensity score method<sup>27</sup>.** A propensity score is the probability of missing data and is estimated by the logistic regression or other methods using observed data. Let  $p_j$  be the missing probability of variable  $X_j$ . For example, in the case of logistic regression, the following model is used for estimation.

$$\text{logit}(p_j) = \mathbf{X}_{-j}^t \boldsymbol{\beta}_{\text{PS}}$$

The data sorted by the estimated propensity scores  $\hat{p}_j$ , are classified into  $k$  strata, and the approximate Bayesian bootstrap method<sup>28</sup> is applied to each stratum. Let  $n_{\text{obs},s}$  be the number of cases observed in the  $s$ th stratum ( $s = 1, 2, \dots, k$ ). A dataset is created by sampling  $n_{\text{obs},s}$  cases with replacement from each stratum of the observed data. The missing values are randomly imputed from the created dataset. In our study, propensity scores were estimated by the logistic regression and  $k = 5$ . Although this method is typically not used in clinical studies with baseline intervention and multiple follow-up data, we used it in the MRT since the interventions differed from time to time, and each time point could be considered separate data.

## Research ethics and patient consent

This research was a simulation study; no data were obtained from new patients. In the application section, we used the data in the Ally Micro-Randomized Trial. All individual participants enrolled in the study provided informed consent.<sup>8</sup> The anonymized dataset was available online, and we used the data after obtaining the permission from the authors.

## Simulation study

### Data generation

We assumed an MRT with 100 participants, 30 time points, continuous outcomes, and three background factors. The simulation of the assumed MRT was repeated 1000 times. Let  $X_{1i}$ ,  $X_{2i}$  and  $X_{3i}$  be the background factors of the  $i$ th participant and let  $A_{it} \in \{0, 1\}$  be the intervention at time  $t$ . The intervention was assigned with a probability of 0.5. The outcome after the intervention at  $t$  was denoted as  $Y_{it+1}$ ; the indicator variable  $R_{it}$  took the value 1 when  $Y_{it}$  was missing and 0 when  $Y_{it}$  was observed; the missing probability of  $Y_{it}$  was denoted as  $P_{it}$ .

For the background factors, two cases were considered: (a) the factors followed the standard normal distribution, and (b) the factors took the value of either 1 or  $-1$  with a probability of 0.5 independently. The outcomes were calculated using the following equation:

$$Y_{it+1} = aA_{it} + \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \beta_3 X_{3i} + \beta_{12} X_{2i} * A_{it} + \varepsilon_{it+1}$$

$\varepsilon_i = (\varepsilon_{i1}, \varepsilon_{i2}, \dots, \varepsilon_{i30})^T$  was the error vector for subject  $i$ .  $\varepsilon_i$  was generated from a multivariate normal distribution with the following covariance matrix:

$$\sigma^2 \begin{pmatrix} 1 & \rho & \rho^2 & \dots & \rho^{29} \\ & 1 & \rho & \dots & \rho^{28} \\ & & 1 & \dots & \rho^{27} \\ & & & \ddots & \vdots \\ & & & & 1 \end{pmatrix}$$

The missing probability was calculated using the following equation:

$$P(R_{it+1} = 1) = \text{logit}^{-1}(\theta A_{it} + \eta_0 + \eta_1 X_{1i} + \eta_2 X_{2i} + \eta_3 X_{3i})$$

Two scenarios with and without the interaction were simulated ( $\beta_{12} = 0$  or  $1.0$ ). The influence of covariates on missing probabilities was simulated in cases with large effects ( $\eta_1 = 0.5$ ,  $\eta_2 = 1$ ,  $\eta_3 = 1.5$ ) and small effects ( $\eta_1 = 0.25$ ,  $\eta_2 = 0.5$ ,  $\eta_3 = 0.75$ ). Other parameters were set as  $a = 0.5$ ,  $\beta_0 = 0$ ,  $\beta_1 = 0.5$ ,  $\beta_2 = 1.0$ ,  $\beta_3 = 1.5$ ,  $\sigma^2 = 1$ ,  $\rho = 0.5$ ,  $\theta = -1.0$ ,  $\eta_0 = -1.0$ .

### Simulation statistical analysis

The generated datasets were analyzed by GEE and RE with individual-level intercept. Covariates were not adjusted in

either analysis. We specify independence for the correlation structure of GEE and used the Kenward–Roger method to estimate the degrees of freedom in RE. We compared the following methods for handling missing data. Multiple imputations were performed within each subset of time points (we called this “by time point”), and multiple imputations were performed on pooled data without distinguishing time points (we called this “disregarding time points”). In the “by time point” method, imputation was performed 30 times on the dataset containing 100 participants, and in the “disregarding time points” method, imputation was performed once on the dataset containing  $100 \times 30$  values. As a reference, we also analyzed the dataset before generating the missing data (Full).

(i) Available data analysis (AD).

If  $X_1$ ,  $X_2$  and  $X_3$  could be accounted for (simulation patterns (ii) to (v)), we considered these patterns as MAR.

(ii) Regression method with  $A$ ,  $X_1$ ,  $X_2$ ,  $X_3$  by time points (Reg\_ByT).

(iii) Regression method with  $A$ ,  $X_1$ ,  $X_2$ ,  $X_3$ , disregarding time points (Reg\_Pool).

(iv) Propensity score method with  $A$ ,  $X_1$ ,  $X_2$ ,  $X_3$  by time points (PS\_ByT).

(v) Propensity score method with  $A$ ,  $X_1$ ,  $X_2$ ,  $X_3$ , disregarding time points (PS\_Pool).

If we hypothetically assume that only  $X_1$  was observed (simulation patterns (vi) to (x)), these patterns could be considered MNAR. We also evaluated the performance of the analysis using random effects in addition to  $X_1$ .

(vi) Regression method with  $A$ ,  $X_1$  by time points (Reg\_ByT).

(vii) Regression method with  $A$ ,  $X_1$ , disregarding time points (Reg\_Pool).

(viii) Regression method with  $A$ ,  $X_1$  and the individual-level random effect as the intercept (Reg\_Re).

(ix) Propensity score method with  $A$ ,  $X_1$  by time points (PS\_ByT).

(x) Propensity score method with  $A$ ,  $X_1$ , disregarding time points (PS\_Pool).

In all multiple imputation, we generated 100 pseudo-datasets and combined the results. We obtained the final results by averaging the point estimates and standard errors of 1000 simulations.

## Simulation results

A consistent pattern was observed in all scenarios, irrespective of whether the covariates were continuous or categorical. Therefore, we present the results without distinguishing between different distributions of the covariates. The results were summarized in four sections based on

presence of the interaction effect between the intervention and the background factor, and whether we accounted for all covariates or only one covariate (MAR or MNAR).

**Under MAR and without interaction.** In the available data analysis, the GEE estimations were biased, but this bias was minimal in the RE estimates (Figure 2(a) and (b)). In the regression methods with  $A$ ,  $X_1$ ,  $X_2$  and  $X_3$ , we obtained unbiased results regardless of whether the time points were disregarded. By contrast, the propensity score method yielded bias in the direction of underestimation. The propensity score method, which disregarded time points, did not obtain some results because of the absence of observed values in the stratified subgroups. The results were obtained 802 times for binary covariates and 862 times for continuous covariates when the effects of the covariates on missingness were large. These results were similar, regardless of the degree to which the covariates affected missingness.

**Under MNAR and without interaction.** In the regression methods with  $A$  and  $X_1$ , we obtained biased results, regardless of whether the time points were disregarded (Figure 3(a) and (b)). The bias was reduced by adding individual-level random effects as intercepts to the imputation model. The propensity score method resulted in bias; however, this bias was smaller than that in the regression model without individual-level random effects. These results were similar, regardless of the degree to which the covariates affected missingness.

**Under MAR and with interaction.** In the available data analysis, both the estimations of GEE and RE were biased (Figure 4(a) and (b)). The regression methods with  $A$ ,  $X_1$ ,  $X_2$  and  $X_3$  resulted in a similar degree of bias as the available data analyses, regardless of whether time points were disregarded. The propensity score method resulted in a more considerable bias. The propensity score method, which disregarded time points, did not obtain some results because of the absence of observed values in the stratified subgroups. The results were obtained 802 times for binary covariates and 862 times for continuous covariates when the effects of the covariates on missingness were large. These results were similar, regardless of the degree to which the covariates affected missingness.

**Under MNAR and with interaction.** In the regression methods with  $A$  and  $X_1$ , we obtained biased results, regardless of whether the time points were disregarded (Figure 5(a) and (b)). The bias was reduced by adding individual-level random effects to the imputation model as intercepts. The propensity score method improved over the available data analysis but resulted in bias.

**Standard errors.** The results regarding standard errors were presented below, without distinguishing between different

mechanisms of missing data, presence of interactions, and degrees of impact of covariates on missingness. In all analyses, standard errors of the RE were smaller than those of GEE. The multiple imputations in GEE outperformed the available data analysis in terms of accuracy, but the multiple imputation in RE had worse accuracy.

## Application study

### *The ally micro-randomized trial (the ally trial)*

We compared the available data analysis and the multiple imputation by applying them to the Ally MRT data. The MRT of the Ally, an app that promotes physical activities in adults, was conducted in Switzerland between October and December in 2017 with 274 health insurance participants. The results of this trial were published, and the anonymized dataset was available online (cited 2023 Apr 13).<sup>8</sup> The purpose of the trial was to evaluate the effects and interactions between the three interventions.

1. Incentive conditions (cash, charity, or no incentives)
2. Weekly plannings (action planning, coping planning, or no planning)
3. Daily self-monitoring prompt (receiving or not)

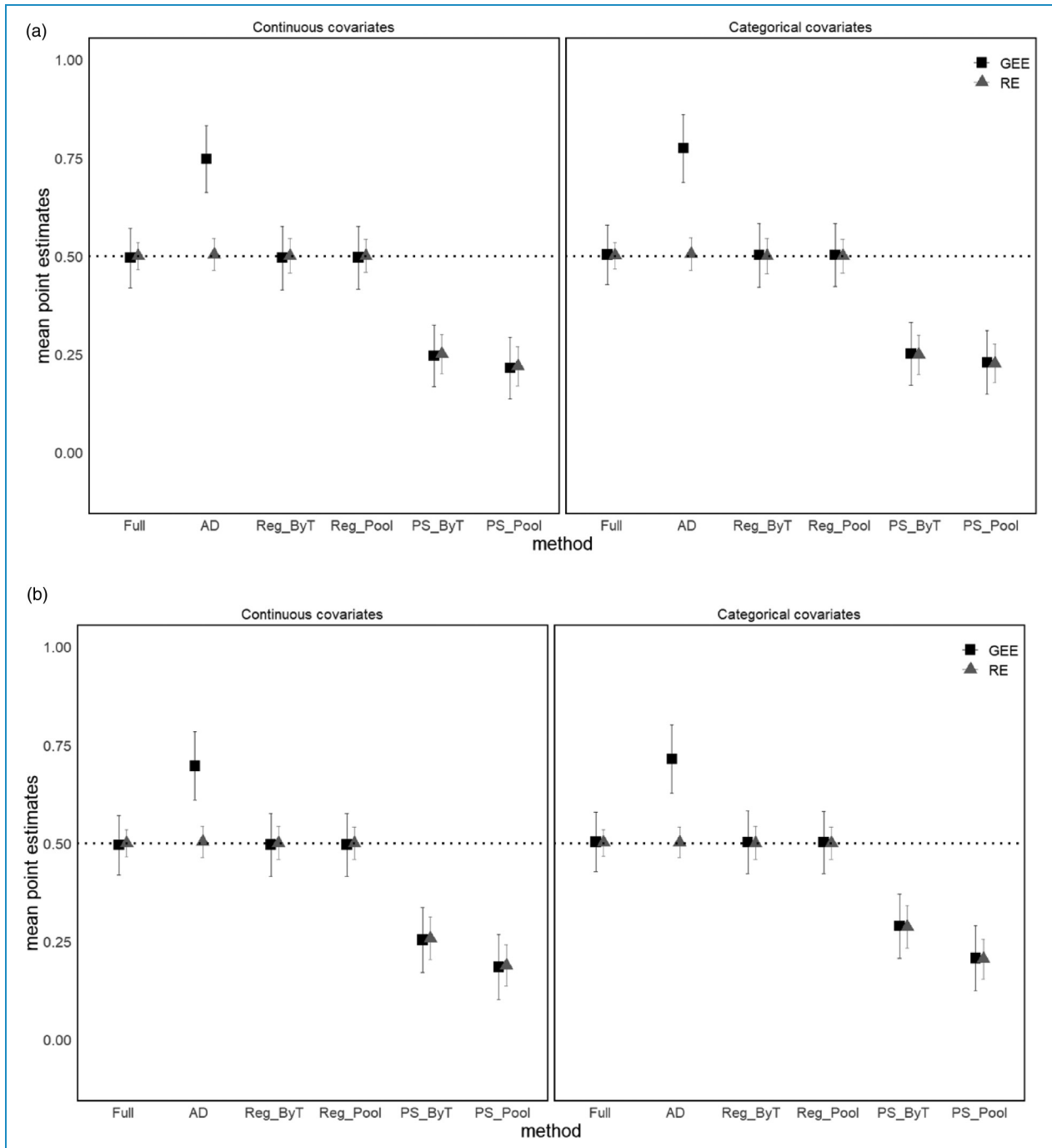
Interaction effects between incentives and self-monitoring prompts (SMPs) and interaction effects between incentives and weekly plannings were postulated when the study was designed.<sup>29</sup> Each intervention was randomized at different times. The SMPs were randomly assigned every day except Sunday, which was micro-randomized. The primary endpoint was achievement of the target number of steps per day, whereas the secondary endpoint was the number of steps taken. The expected number of steps was 11508, calculated as 274 participants for 42 days. However, 34.4% (3957/11508) of the records were missing. The published paper reported that only incentives were effective.<sup>8</sup>

### *Application statistical analysis*

We estimated the step ratios and 95% confidence intervals (CIs) for the three interventions, the interactions between incentives and plannings, and the interactions between incentives and SMPs. The log-transformed numbers of steps were analyzed as the outcomes using GEE and RE. We specified independence for the correlation structure of GEE and used the Kenward–Roger method to estimate the degrees of freedom in RE. In RE, to account for the nesting of interventions (split-split-plot design), we specified the interactions between participants and incentives, and the interactions between participants, incentives, weeks, and plannings as random effects.

The available dataset included age, gender, employment status, smartphone operating system, and the number of steps taken during the baseline period as background



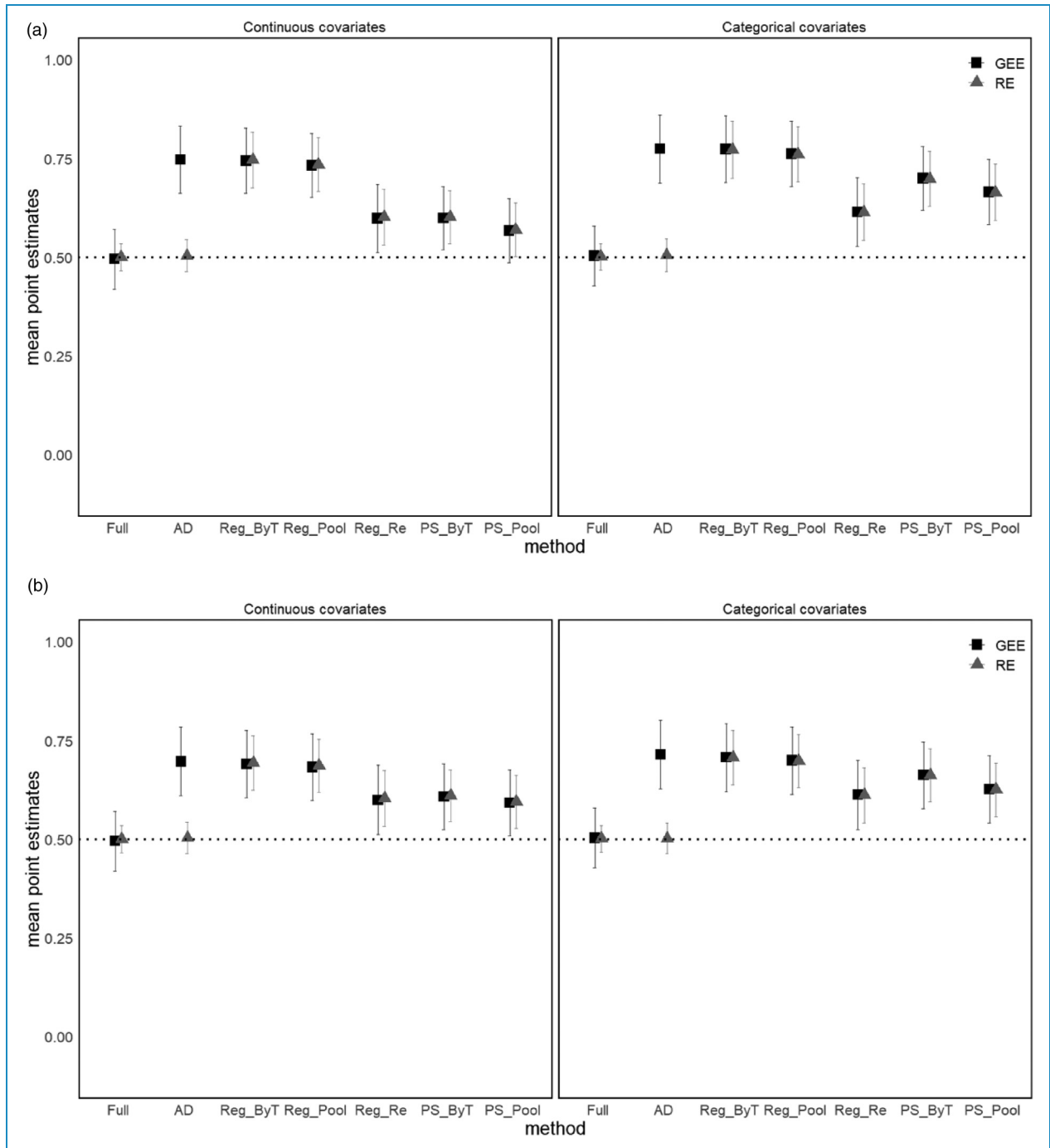


**Figure 2.** Simulation results of mean point estimates  $\pm$  average standard errors when all covariates were accounted for, there was no interaction, and the effects of covariates on missingness were (a) large and (b) small.

factors. The number of steps taken was imputed from these background factors and interventions at each time point. Because there were missing background factors, we imputed the background factors with multiple imputation by chained equations (MICE)<sup>30</sup> among background factors and then imputed the missing step counts. In MICE, categorical variables were imputed with the logistic

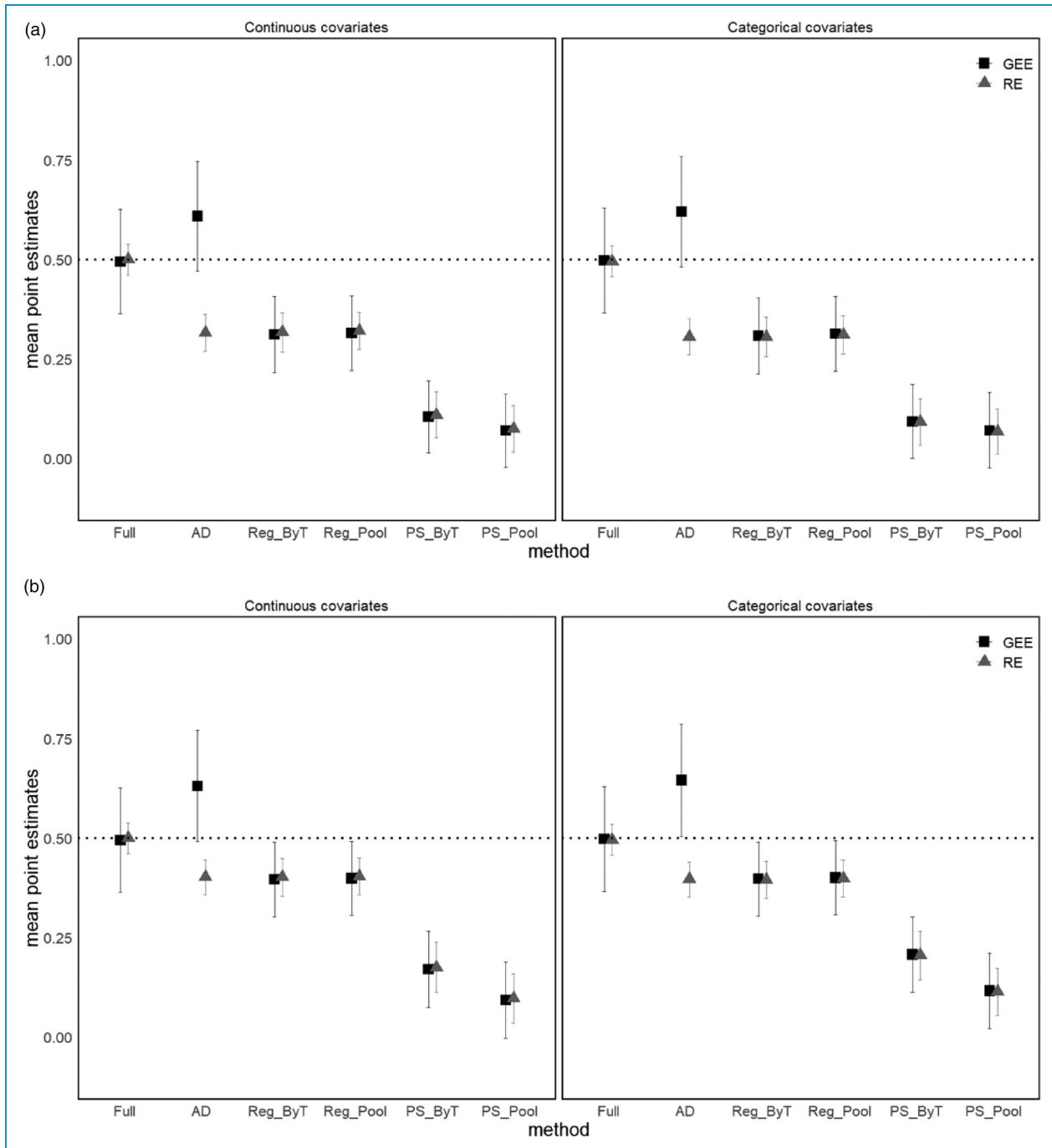
regression, and continuous variables were imputed with the predictive mean matching.<sup>31,32</sup> We compared the following methods for handling missing step counts.

- (i) Available data analysis.
- (ii) Regression method with interventions and background factors by time points.



**Figure 3.** Simulation results of mean point estimates  $\pm$  average standard errors when only one covariate was accounted for, there was no interaction, and the effects of covariates on missingness were (a) large and (b) small. Values in parentheses are average standard errors.

- (iii) Regression method with interventions and background factors, disregarding time points.
  - (iv) Regression method with interventions, background factors and the individual-level random effect as the intercept.
  - (v) Propensity score method with interventions and background factors by time points.
  - (vi) Propensity score method with interventions and background factors, disregarding time points.
- In all multiple imputation, we generated 100 pseudo-datasets and combined the results. SAS 9.4 was used for all statistical analyses.



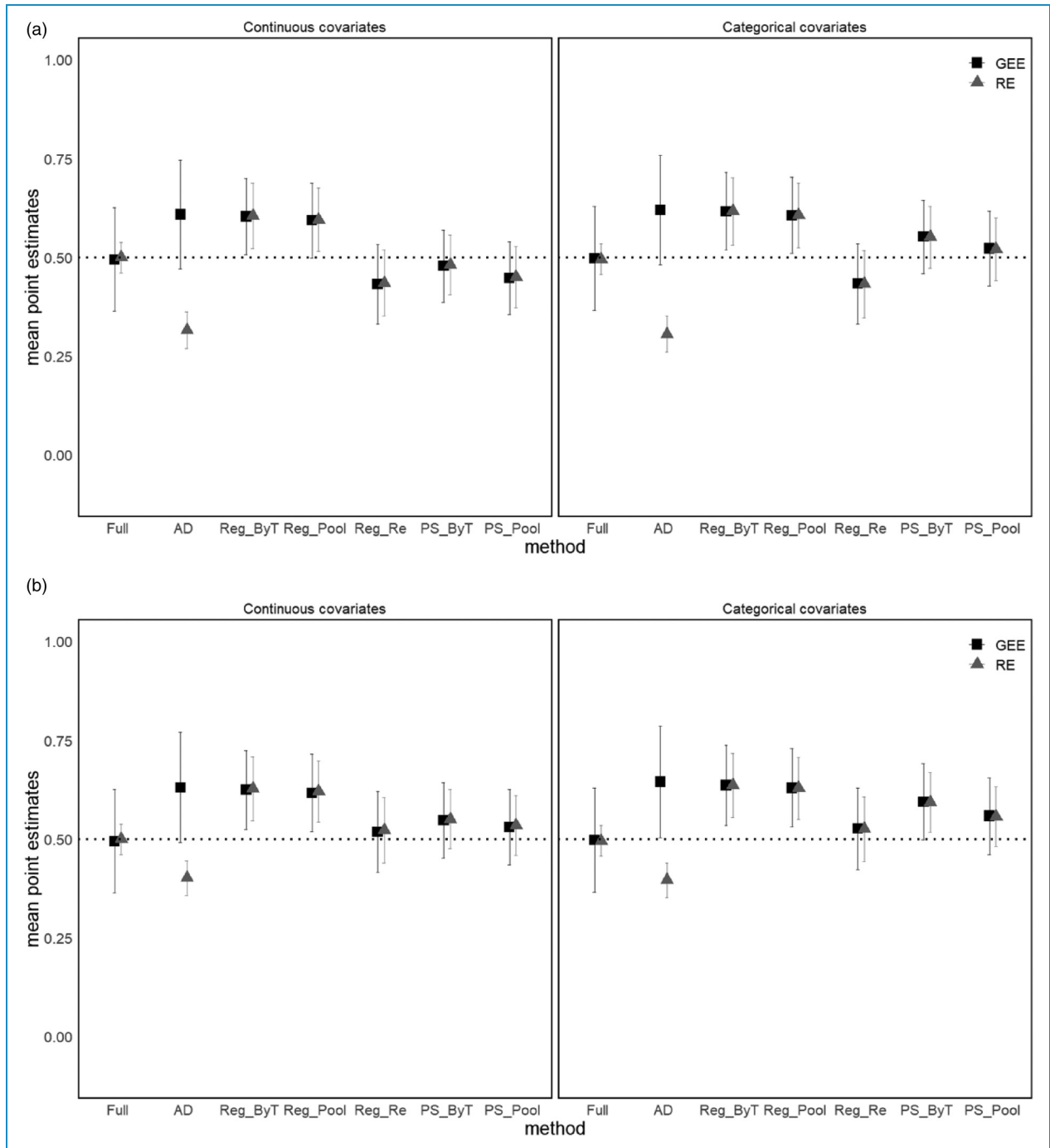
**Figure 4.** Simulation results of mean point estimates  $\pm$  average standard errors when all covariates were accounted for, there was the interaction, and the effects of covariates on missingness were (a) large and (b) small.

### Application results

The estimates of the step ratio for the SMPs, which were micro-randomized in the study, remained consistent among different methods for handling missing data (Table 1). Other estimates did not also change considerably (Table 2, Table 3).

Except for incentive interventions, the point estimates of the step ratios ranged from 0.9 to 1.1, and the 95% CIs included 1.0. The point estimate of the step ratio for the incentive intervention was approximately 1.1. In some multiple imputation, the lower limits of the 95% CIs exceeded 1.0. GEE resulted in narrower CIs for the





**Figure 5.** Simulation results of mean point estimates  $\pm$  average standard errors when only one covariate was accounted for, there was the interaction, and the effects of covariates on missingness were (a) large and (b) small.

incentive interventions and wider CIs for the SMPs compared to RE.

**Discussion**

*Principal findings*

We evaluated the performance of the available data analysis and the multiple imputation in MRTs in which interventions

were uniformly randomized for all participants through the trial and without considering time-dependent covariates in the analysis. We considered 16 MRT scenarios in simulation based on the presence of the interaction effect between the background factor and the intervention, the presence of unmeasured background factors, the distributions of background factors, and the influence of covariates on the missing probabilities. The three main findings were

**Table 1.** Estimates and 95% CI of the step ratios of self-monitoring prompts for the Ally trial.

Method for handling missing data	GEE		RE	
	Estimate	95% CI	Estimate	95% CI
Available data analysis	1.000	[0.953–1.049]	0.993	[0.954–1.034]
Regression method by time points	1.000	[0.950–1.052]	1.000	[0.956–1.046]
Regression method, disregarding time points	0.998	[0.950–1.049]	0.998	[0.956–1.042]
Regression method with random effects as intercepts	1.000	[0.951–1.052]	1.001	[0.957–1.046]
Propensity score method by time points	1.007	[0.957–1.059]	1.006	[0.961–1.054]
Propensity score method, disregarding time points	0.999	[0.952–1.049]	0.999	[0.956–1.043]

CI, confidence interval; Ally, assistant to lift your level of activity; SMP, self-monitoring prompt; GEE, generalized estimating equations; RE, random effects model.

as follows: first, the available data analysis could be an option for handling missing data; second, including random effects in the regression methods reduced bias; and third, the propensity score method performed poorly, even when the mechanism of missing data was MAR.

Our first main finding was that there were some scenarios in which the available data analysis performed better than the multiple imputation. In general, bias in the available data analysis can occur because of unmeasured factors influencing both missing data and outcomes.<sup>23</sup> In our study, bias was observed in GEE. However, when using RE for the available data analysis in scenarios without the interaction effect, the bias was minimal. In the presence of time-dependent covariates, RE is not considered valid, and methods based on estimating equations have been proposed.<sup>5,22,33</sup> However, within the MRTs we focused on, our results suggested that RE may be better than GEE. These results are consistent with the statements in Little<sup>14</sup>: “If the data are MAR, likelihood inference based on the random effects model is equivalent to inference based on the full likelihood considering missing-data mechanism. In contrast, GEE generally requires a stronger assumption to yield a consistent estimator”. Nonetheless, these results do not necessarily imply that the available data analysis by RE is superior to other methods. In our study, bias was evident in scenarios with an interaction between the covariate and the intervention, similar to observations in GEE. This could be because the random effects partially reflected the effects of the unobserved covariates in the patterns without the interaction, and the random effects corresponding to the interaction were not included in the patterns with the interaction.

Our second key finding was that including random effects in the regression model of the multiple imputation could help reduce bias. As expected, the multiple imputation with the

regression model performed well only in scenarios where we accounted for all covariates and without the interaction, whereas bias was observed in scenarios where we accounted for only one covariate or with interaction effects. However, the bias was reduced by including individual-level random effects as intercepts in the imputation models with only one covariate. The results indicate that including individual-level random effects could reduce the influence of unmeasured background factors.

Our final key finding was that the propensity score method could introduce bias, even in situations where the missing probability model was correct. The propensity score method does not consider the association between the variables used to estimate the probabilities of missing data and the outcome variable. When the intervention affects the outcome variable but not the missingness, the intervention effect is underestimated because the associations between the outcome variable and the interventions are not correctly reflected in the imputations.<sup>34</sup> In the simulations, the impact of background factors on missing data was greater than that of the intervention, and the effect of the intervention on the outcome may have been underestimated due to the relatively small impact of the intervention on missingness. When imputations were performed at each time point, the observed subjects in the subgroups stratified by missing probabilities were absent and could not be completed. In MRTs with a small number of participants, estimates may not be obtained if missing data are handled using the propensity score method at each time point.

In our study, multiple imputation had shown better efficiency when we used GEE but not when we used RE. This comparison conclusions of efficiency in RE align with previous studies.<sup>35–37</sup>

Based on the findings of this study, it is challenging to determine the best method of handling missing data in

Table 2. Estimates and 95%CI of the step ratio for the Ally trial by GEE.

Variable	Step ratio (95%CI)					
	Available data	MI 1 <sup>a</sup>	MI 2 <sup>b</sup>	MI 3 <sup>c</sup>	MI 4 <sup>d</sup>	MI 5 <sup>e</sup>
Cash incentive	1.132 (0.978–1.310)	1.150 (1.066–1.241)	1.142 (1.060–1.231)	1.144 (1.055–1.241)	1.099 (1.024–1.179)	1.102 (1.024–1.185)
Charity incentive	1.118 (0.970–1.290)	1.139 (1.059–1.226)	1.130 (1.048–1.218)	1.120 (1.030–1.217)	1.077 (1.005–1.153)	1.075 (1.002–1.153)
AP	0.985 (0.917–1.058)	1.012 (0.951–1.077)	1.000 (0.941–1.062)	1.004 (0.944–1.068)	0.998 (0.940–1.060)	0.991 (0.930–1.056)
CP	0.997 (0.936–1.063)	0.998 (0.938–1.063)	0.992 (0.933–1.055)	0.998 (0.938–1.061)	1.005 (0.946–1.068)	0.999 (0.941–1.060)
Cash * AP	0.993 (0.901–1.096)	0.988 (0.907–1.076)	0.997 (0.919–1.082)	0.999 (0.916–1.088)	0.993 (0.911–1.082)	0.992 (0.909–1.083)
Cash * CP	0.980 (0.889–1.079)	0.991 (0.911–1.079)	0.992 (0.912–1.078)	0.992 (0.913–1.078)	0.982 (0.902–1.068)	0.982 (0.902–1.070)
Charity * AP	1.076 (0.974–1.188)	1.024 (0.941–1.115)	1.038 (0.958–1.126)	1.035 (0.951–1.127)	1.042 (0.959–1.133)	1.048 (0.962–1.142)
Charity * CP	0.971 (0.886–1.063)	0.972 (0.892–1.059)	0.976 (0.899–1.059)	0.975 (0.897–1.059)	0.982 (0.902–1.069)	0.982 (0.905–1.066)
SMP	1.000 (0.953–1.049)	1.000 (0.950–1.052)	0.998 (0.950–1.049)	1.000 (0.951–1.052)	1.007 (0.957–1.059)	0.999 (0.952–1.049)
Cash * SMP	1.021 (0.960–1.085)	1.016 (0.948–1.089)	1.016 (0.949–1.087)	1.015 (0.948–1.087)	1.014 (0.944–1.089)	1.015 (0.949–1.086)
Charity * SMP	0.994 (0.933–1.059)	0.997 (0.931–1.067)	0.995 (0.929–1.065)	0.995 (0.930–1.065)	0.994 (0.926–1.066)	0.996 (0.931–1.066)

CI, confidence interval; Ally, assistant to lift your level of activity; GEE, generalized estimating equations; AP, action planning; CP, coping planning; SMP, self-monitoring prompt

<sup>a</sup>Regression method with interventions and backgrounds by time point.

<sup>b</sup>Regression method with interventions and backgrounds, disregarding time points.

<sup>c</sup>Regression method with interventions, backgrounds and the individual-level random effects as intercept.

<sup>d</sup>Propensity score method with interventions and backgrounds by time point.

<sup>e</sup>Propensity score method with interventions and backgrounds, disregarding time points.

Table 3. Estimates and 95%CI of the step ratio for the Ally trial by the random effects model.

Step ratio (95%CI)						
Variable	Available data	MI 1 <sup>a</sup>	MI 2 <sup>b</sup>	MI 3 <sup>c</sup>	MI 4 <sup>d</sup>	MI 5 <sup>e</sup>
Cash incentive	1.111 (0.952–1.296)	1.150 (1.014–1.303)	1.141 (1.007–1.293)	1.144 (1.004–1.302)	1.098 (0.981–1.229)	1.101 (0.982–1.234)
Charity incentive	1.067 (0.915–1.244)	1.139 (1.008–1.286)	1.129 (0.998–1.278)	1.120 (0.983–1.274)	1.076 (0.964–1.202)	1.074 (0.961–1.201)
AP	1.005 (0.945–1.067)	1.012 (0.953–1.075)	1.000 (0.944–1.059)	1.004 (0.947–1.065)	0.998 (0.940–1.060)	0.991 (0.930–1.056)
CP	1.005 (0.945–1.068)	0.998 (0.940–1.060)	0.992 (0.935–1.052)	0.998 (0.941–1.058)	1.005 (0.946–1.068)	0.999 (0.941–1.060)
Cash * AP	1.002 (0.922–1.089)	0.988 (0.910–1.072)	0.997 (0.923–1.077)	0.999 (0.920–1.084)	0.993 (0.911–1.082)	0.992 (0.909–1.083)
Cash * CP	0.997 (0.917–1.084)	0.991 (0.914–1.075)	0.992 (0.916–1.074)	0.992 (0.917–1.074)	0.982 (0.902–1.068)	0.982 (0.902–1.069)
Charity * AP	1.056 (0.969–1.150)	1.024 (0.944–1.111)	1.038 (0.962–1.121)	1.035 (0.955–1.122)	1.042 (0.958–1.133)	1.048 (0.962–1.142)
Charity * CP	0.972 (0.892–1.059)	0.972 (0.896–1.055)	0.976 (0.903–1.055)	0.975 (0.901–1.055)	0.982 (0.901–1.069)	0.982 (0.905–1.066)
SMP	0.993 (0.954–1.034)	1.000 (0.956–1.046)	0.998 (0.956–1.042)	1.001 (0.957–1.046)	1.006 (0.961–1.054)	0.999 (0.956–1.043)
Cash * SMP	1.029 (0.974–1.087)	1.018 (0.958–1.081)	1.017 (0.959–1.078)	1.016 (0.958–1.078)	1.016 (0.952–1.085)	1.017 (0.957–1.080)
Charity * SMP	1.006 (0.951–1.064)	0.997 (0.940–1.058)	0.995 (0.937–1.057)	0.996 (0.940–1.056)	0.995 (0.933–1.060)	0.997 (0.938–1.059)

CI, confidence interval; Ally, assistant to lift your level of activity; AP, action planning; CP, coping planning; SMP, self-monitoring prompt.

<sup>a</sup>Regression method with interventions and backgrounds by time point.

<sup>b</sup>Regression method with interventions and backgrounds, disregarding time points.

<sup>c</sup>Regression method with interventions, backgrounds and the individual-level random effects as intercept.

<sup>d</sup>Propensity score method with interventions and backgrounds by time point.

<sup>e</sup>Propensity score method with interventions and backgrounds, disregarding time points.

MRTs. However, we can propose several recommendations. The available data analysis of the RE might perform others when interaction effects are not considered. Nonetheless, this assumption may not hold true in the context of MRTs. Within the scope of our study, we found that multiple imputation by regression methods, including individual-level random effects, yielded better results under such conditions.

### Limitations and future research

This study had some limitations. First, it assumed that background factors and the intervention determined the presence of missingness. In the Ally trial, since the number of steps was measured automatically, the association between missingness and previous outcomes was expected to be weak. However, in other MRTs, there may be a strong association between missingness and previous outcomes. Therefore, it is necessary to research cases in which missingness depends on the previous outcomes in the future. In MRTs with non-monotonic missing data, multiple imputation conditioned on variables of all time points is unrealistic owing to overfitting and problems with the number of parameters. One possible approach would be to apply a two-fold FCS<sup>38</sup> that performs multiple imputation for separate time points.

Second, time-varying covariates were not considered in this study. The time-varying variables analyzed in the Ally trial were the interventions and the outcome, and only the outcome could be missing. However, time-dependent covariates are collectible using wearable devices to optimize interventions.<sup>4,39</sup> The analysis of such app studies should consider coping with time-varying covariates; hence, this is an important topic to investigate in the future. In MRTs where time-dependent covariates are used for only statistical power, sensitivity analysis should be performed using missing handling methods without including time-varying covariates in the analysis.

Third, this study examined the performance of the multiple imputation in situations where the outcome was a continuous variable but not in situations where the outcome was a categorical variable.

Fourth, we compared the performance of traditional multiple imputation methods. Multiple imputation using machine learning and other methods have been developed,<sup>40–42</sup> and these methods can potentially implement multiple imputation methods with fewer assumptions. The evaluation of applying these methods to the missing data in MRTs is one of the important future works.

### Application discussion

In the Ally trial, the expected number of records was 11508, calculated as 274 participants for 42 days, but 34.4% (3957/11508) of records of daily steps were missing. Although

many values were missing, the results were similar between the approaches. It was suggested that the study might receive little influence owing to missing data. We found no significant intervention effects in the available data analysis. The paper reporting the results of the Ally trial<sup>8</sup> also found no significant effect when the number of steps was analyzed as a continuous quantity, and the results in this study were considered consistent.

GEE tended to obtain narrower CIs for incentives and wider CIs for SMPs than RE. This result appropriately reflects the study design of split-split plot design. RE that appropriately reflects the study design should be selected.

### Conclusion

We compared the performance of multiple imputation and the available data analysis in simple MRTs, where missing data depend on the intervention and background factors. The available data analysis of the random-effects model performed better if there was no interaction between the intervention and the background factor. However, we recommend using the multiple imputation when performing GEE or when interactions are assumed. In addition, the imputation model in regression methods should include individual-level random effects. When using propensity score methods, it is advisable to exclude variables with minimal influence on the outcome.

### Abbreviations

Ally	Assistant to Lift your Level of activity
CI	confidence interval
FCS	fully conditional specification
GEE	generalized estimating equations
MAR	missing at random
MCAR	missing completely at random
MICE	multiple imputation by chained equations
MNAR	missing not at random
MRT	micro-randomized trial
RE	random effects model
SMP	self-monitoring prompt

**Acknowledgements:** The authors acknowledge the contributions of the researchers who conducted the MRT of Ally and provided data to the public.

**Contributors:** MK and KO were involved in research planning, data analysis, and writing the manuscript. All authors reviewed and edited the manuscript and approved the final version of the manuscript.

**Declaration of conflicting interests:** The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

**Ethical approval:** Not applicable.

**Funding:** The authors disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This research was supported by AMED under Grant Number JP211k0201701 and the National Cancer Center Research and Development Fund (2021-A-12 to H.B.).

**Guarantor:** KO

**ORCID ID:** Masahiro Kondo  <https://orcid.org/0000-0003-4647-3746>

## References

- Bates DW, Landman A and Levine DM. Health apps and health policy: what is needed? *JAMA* 2018; 320: 1975–1976.
- Gordon WJ, Landman A, Zhang H, et al. Beyond validation: getting health apps into clinical practice. *NPJ Digit Med* 2020; 3: 14.
- Direito A, Dale LP, Shields E, et al. Do physical activity and dietary smartphone applications incorporate evidence-based behaviour change techniques? *BMC Public Health* 2014; 14: 646.
- Klasnja P, Hekler EB, Shiffman S, et al. Microrandomized trials: an experimental design for developing just-in-time adaptive interventions. *Health Psychol* 2015; 34: 1220–1228.
- Qian T, Walton AE, Collins LM, et al. The microrandomized trial for developing digital interventions: experimental design and data analysis considerations. *Psychol Methods* 2022; 27(5): 874–894.
- Aguilera A, Hernandez-Ramos R, Haro-Ramos AY, et al. A text messaging intervention (StayWell at home) to counteract depression and anxiety during COVID-19 social distancing: pre-post study. *JMIR Ment Health* 2021; 8: e25298.
- Figueroa CA, Deliu N, Chakraborty B, et al. Daily motivational text messages to promote physical activity in university students: results from a microrandomized trial. *Ann Behav Med* 2022; 56: 212–218.
- Kramer J-N, Künzler F, Mishra V, et al. Which components of a smartphone walking app help users to reach personalized step goals? Results from an optimization trial. *Ann Behav Med* 2020; 54: 518–528.
- Nordby ES, Gjestad R, Kenter RMF, et al. The effect of SMS reminders on adherence in a self-guided internet-delivered intervention for adults with ADHD. *Front Digit Health* 2022; 4: 821031.
- Militello L, Sobolev M, Okeke F, et al. Digital prompts to increase engagement with the Headspace app and for stress regulation among parents: feasibility study. *JMIR Form Res* 2022; 6: e30606.
- Klasnja P, Smith S, Seewald NJ, et al. Efficacy of contextually tailored suggestions for physical activity: a micro-randomized optimization trial of HeartSteps. *Ann Behav Med* 2019; 53: 573–582.
- Meyerowitz-Katz G, Ravi S, Arnolda L, et al. Rates of attrition and dropout in app-based interventions for chronic disease: systematic review and meta-analysis. *J Med Internet Res* 2020; 22: e20283.
- Rubin DB. Inference and missing data. *Biometrika* 1976; 63: 581–592.
- Little RJA. Modeling the drop-out mechanism in repeated-measures studies. *J Am Stat Assoc* 1995; 90: 1112.
- Robins JM, Rotnitzky A and Zhao LP. Analysis of semi-parametric regression models for repeated outcomes in the presence of missing data. *J Am Stat Assoc* 1995; 90: 106–121.
- Little RJ, D’Agostino R, Cohen ML, et al. The prevention and treatment of missing data in clinical trials. *N Engl J Med* 2012; 367: 1355–1360.
- Curnow E, Carpenter JR, Heron JE, et al. Multiple imputation of missing data under missing at random: compatible imputation models are not sufficient to avoid bias if they are mis-specified. *J Clin Epidemiol* 2023; 160: 100–109.
- Vollrath ME and Torgersen S. Personality types and risky health behaviors in Norwegian students. *Scand J Psychol* 2008; 49: 287–292.
- Rhodes RE and Smith NEI. Personality correlates of physical activity: a review and meta-analysis. *Br J Sports Med* 2006; 40: 958–965.
- Pampel FC, Krueger PM and Denney JT. Socioeconomic disparities in health behaviors. *Annu Rev Sociol* 2010; 36: 349–370.
- Albert PS. Longitudinal data analysis (repeated measures) in clinical trials. *Stat Med* 1999; 18: 1707–1732.
- Qian T, Klasnja P and Murphy SA. Linear mixed models with endogenous covariates: modeling sequential treatment effects with application to a mobile health study. *Stat Sci* 2020; 35: 375–390.
- National Research Council (US) Panel on Handling Missing Data in Clinical Trials. *The prevention and treatment of missing data in clinical trials*. Washington, DC: National Academies Press (US), 2014.
- Rubin DB. Multiple imputations in sample surveys—a phenomenological Bayesian approach to nonresponse.
- van Buuren S. Multiple imputation of discrete and continuous data by fully conditional specification. *Stat Methods Med Res* 2007; 16: 219–242.
- Rubin DB. *Multiple imputation for nonresponse in surveys*. New York: John Wiley & Sons, 1987.
- Lavori PW, Dawson R and Shera D. A multiple imputation strategy for clinical trials with truncation of patient data. *Stat Med* 1995; 14: 1913–1925.
- Rubin DB and Schenker N. Multiple imputation for interval estimation from simple random samples with ignorable nonresponse. *J Am Stat Assoc* 1986; 81: 366–374.
- Kramer J-N, Künzler F, Mishra V, et al. Investigating intervention components and exploring states of receptivity for a smartphone app to promote physical activity: protocol of a microrandomized trial. *JMIR Res Protoc* 2019; 8: e11540.
- Buuren S and Oudshoorn C. Multivariate imputation by chained equations: Mice V1.0 user’s manual. <https://www.semanticscholar.org/paper/015d352b1c71acfacaca59377d524a1f35245244> (2000, accessed 1 December 2021).
- Little RJA. Missing-data adjustments in large surveys. *J Bus Econ Stat* 1988; 6: 287–296.
- Rubin DB. Statistical matching using file concatenation with adjusted weights and multiple imputations. *J Bus Econ Stat* 1986; 4: 87–94.



33. Boruvka A, Almirall D, Witkiewitz K, et al. Assessing time-varying causal effect moderation in mobile health. *J Am Stat Assoc* 2018; 113: 1112–1121.
  34. Schafer JL. Multiple imputation: a primer. *Stat Methods Med Res* 1999; 8: 3–15.
  35. Li X, Mehrotra DV and Barnard J. Analysis of incomplete longitudinal binary data using multiple imputation. *Stat Med* 2006; 25: 2107–2124.
  36. Gazel SER and Hayrettin O. Use of generalized estimating equations with multiple imputations for missing longitudinal data. *Yüzüncü Yıl Üniversitesi Fen Bilimleri Enstitüsü Dergisi* 2018; 23: 96–103.
  37. Huque MH, Carlin JB, Simpson JA, et al. A comparison of multiple imputation methods for missing data in longitudinal studies. *BMC Med Res Methodol* 2018; 18: 168.
  38. Nevalainen J, Kenward MG and Virtanen SM. Missing values in longitudinal dietary data: a multiple imputation approach based on a fully conditional specification. *Stat Med* 2009; 28: 3657–3669.
  39. Nahum-Shani I, Smith SN, Spring BJ, et al. Just-in-time adaptive interventions (JITAIs) in mobile health: key components and design principles for ongoing health behavior support. *Ann Behav Med* 2018; 52: 446–462.
  40. Stekhoven DJ and Bühlmann P. Missforest—non-parametric missing value imputation for mixed-type data. *Bioinformatics* 2012; 28: 112–118.
  41. Laqueur HS, Shev AB and Kagawa RMC. SuperMICE: an ensemble machine learning approach to multiple imputation by chained equations. *Am J Epidemiol* 2022; 191: 516–525.
  42. Liu M, Li S, Yuan H, et al. Handling missing values in healthcare data: a systematic review of deep learning-based imputation techniques. *Artif Intell Med* 2023; 142: 102587.
-