

Multiply Robust Weighted Generalized Estimating Equations for Incomplete Longitudinal Binary Data Using Empirical Likelihood

Hiroshi Komazaki^{a,b}, Masaaki Doi^c, Naohiro Yonemoto^{d,e}, Tosiya Sato^a*

^aDepartment of Biostatistics, Kyoto University School of Public Health, Kyoto, Japan

^bClinical Development Department, Maruho Co., Ltd, Kyoto, Japan

^cData Science, Ono Pharmaceutical Co., Ltd., Osaka, Japan

^dDepartment of Neuropsychopharmacology, National Center of Neurology and Psychiatry, Tokyo, Japan

^eDepartment of Public Health, Juntendo University School of Medicine, 2-1-1 Hongo, Bunkyo, Tokyo, Japan.

**Corresponding author: E-mail: komazaki.hiroshi.c48@kyoto-u.jp; Telephone number: +81-75-753-4475; Fax number: +81-75-753-4487*

Multiply Robust Weighted Generalized Estimating Equations for Incomplete Longitudinal Binary Data Using Empirical Likelihood

In clinical trials, missing data may lead to serious misinterpretation of trial results. To address this issue, it is important to collect post-randomization data (such as efficacy measurement data and adverse event onset data). Such post-randomization data are called auxiliary variables and they can be useful for constructing missingness and imputation models. A multiply robust estimator using an empirical likelihood method was previously proposed by Han and Wang (2013) and Han (2014). However, that estimator was developed for cross-sectional data and situations in which no auxiliary variables are missing. This is contrary to actual clinical trial settings, in which some auxiliary variables will invariably be missing. Consequently, to apply Han's method to longitudinal data, missing auxiliary variables need to be imputed. This paper proposes a new method that extends Han's method to a longitudinal outcome model by applying weighted generalized estimating equations with new weights. Monte Carlo simulations of a repeated binary response with missing at random dropouts demonstrated that the proposed estimator is multiply robust and exhibits better performance than that of augmented inverse probability weighted complete-case estimating equations under several simulation scenarios. We also successfully applied the proposed method to plaque psoriasis study data.

Keywords: auxiliary variables; imputation model; missing at random; Monte Carlo simulation; plaque psoriasis

1 Introduction

Consider a randomized, placebo-controlled, double-blind, parallel-group, phase III trial that collects longitudinal data and assesses the drug efficacy at a planned last time point. The estimand is the adjusted odds ratio between the improvement proportions of the treatment groups when all participants complete the study, which corresponds to a parameter of the treatment group in a fitted regression model. In such clinical trials, some of the participants are lost to follow-up, resulting in missing primary endpoints. If the missing data mechanism is missing at random (MAR) or missing not at random, complete-case analysis may lead to a biased estimate of the treatment effect (Little and Rubin 2002).

To address this issue, Han and Wang (2013) and Han (2014) proposed a multiply robust estimator using an empirical likelihood (EL) method (Owen 2001). Han and Wang (2013) and Han (2014) derived a weight based on the EL method and solved the weighted estimating equations to obtain a multiply robust estimator. This method has two advantages. First, it prevents the weight values from being extreme. In the inverse probability weighting (IPW) method (Horvitz and Thomson 1952), the probability of measurements being observed is modeled, and the complete-case data are weighted by the inverse of that probability; however, this method has low estimation efficiency, especially when the estimated probability of being observed is close to zero (i.e., an extreme weight) (Kang and Schafer 2007). Second, the multiply robust

estimator allows multiple missingness and imputation models to mitigate the impact of model misspecification; if a correct model is included among these models, the treatment effect estimator is consistent. In the missingness and imputation models, post-randomization data (e.g., an efficacy measurement and an onset of adverse event) are sometimes included as explanatory variables, i.e., auxiliary variables. These variables are possibly influenced by treatment, correlated with missingness on an outcome and/or outcome itself, and improving efficiency (Mallinckrodt 2013). Han and Wang (2013) and Han (2014) assumed the response variable to be a scalar and that all auxiliary variables are observed. However, in actual clinical settings, some auxiliary variables are usually missing.

In this paper, we propose a new method that extends Han's method to a longitudinal outcome model by applying weighted generalized estimating equations (wGEE) with new weights. The resulting estimator is multiply robust, in the sense that the treatment effect estimator is consistent if any one of the specified missingness or imputation models is correct. Our method accounts for the within-subject correlation structure of the repeated measurement. Thus, the proposed estimator achieves better performance than that of Han in situations where some of the auxiliary variables are missing (i.e., in coarsened longitudinal data analysis).

A few doubly robust methods for application to longitudinal data with dropout have been proposed. Bang and Robins (2005) and Seaman and Copas (2009) proposed

an approach that uses augmented inverse probability weighted complete-case estimating equations (AIPW). The estimator is constructed by the inverse probability weighted GEE and augmentation term imputed with Paik's imputation method (1997), and allows the use of only one missingness model and one imputation model, thus achieving double robustness. Others (Han et al. 2015; Han 2016) also proposed estimators using the EL method, which considers longitudinal data with dropouts. The estimator proposed by Han et al. (2015) is based on the numerical implementation of the conditional EL method, and allows the use of only one missingness model and one imputation model, thus achieving double robustness. The estimator proposed by Han (2016) is based on a calibrating method for missingness probability using data from past visits, and allows the use of only one missingness model and multiple imputation models. When the missingness model is correctly specified, intrinsic efficiency is guaranteed. However, the method targets estimation of a mean of a response at the end of a longitudinal study instead of estimation of a parameter (such as a treatment effect); thus, the estimator most recently proposed by Han (2016) was developed under different conditions compared to Han and Wang (2013), Han (2014), and our proposed method.

The remainder of this paper is organized as follows. In Section 1.1, we describe the plaque psoriasis trial used as a case study. In Section 2, we introduce the proposed method and relevant notation. In Section 3, we outline Monte Carlo simulation studies conducted to evaluate the performance of the proposed method. In Section 4, we

describe the application of the proposed method to the plaque psoriasis study data. Finally, we provide a brief discussion and the conclusions of the study in Section 5.

1.1 Case Study

The efficacy and safety of the study drug M518101 were evaluated in a randomized, placebo-controlled, double-blind, parallel-group, multi-center, phase III study in participants with plaque psoriasis (registered at www.clinicaltrials.gov as NCT01878461). Efficacy assessments were taken at baseline and Weeks 2, 4, 6, and 8. The primary efficacy endpoint was the success rate based on the Investigator Global Assessment (IGA) score at Week 8. The proportions of missing data at Week 8 were 14.5% (53/365) and 13.0% (23/177) in the M518101 and placebo groups, respectively. If, for example, only the participants who showed improvement completed the trial, the complete-case analysis could be biased. In this study, because the estimand was the treatment effect supposing that all subjects adhered to the treatment and completed the study, it was also necessary to consider the influence of the dropout mechanism.

2 Methods

2.1 Notation

Let y_{ij} denote the binary efficacy outcome of a participant $i (= 1, \dots, N)$ at a time point $j (= 1, \dots, T)$, and $\mathbf{Y}_i = (y_{i1}, \dots, y_{iT})^T$, where stochastic independence of y_{i_1j} and y_{i_2j} ,

$i_1 \neq i_2$, is assumed. Let \mathbf{X}_i and R_{ij} denote the design variables (e.g., treatment, baseline covariates) and the indicator variable of observing y_{ij} , respectively. The post-randomization data (i.e., auxiliary variables) are represented as \mathbf{Z}_{ij} , and $\bar{\mathbf{Z}}_{ij} = (\mathbf{Z}_{i1}, \dots, \mathbf{Z}_{ij})^T$. To simplify an explanation, we assume that \mathbf{Z}_{ij} includes only y_{ij} . If both y_{ij} and \mathbf{Z}_{ij} are observed, $R_{ij} = 1$; otherwise, $R_{ij} = 0$. The situation is reasonable because it is common that these data are observed simultaneously at each subject-visit. We assume a monotone missing data pattern and $R_{i1} = 1$. Thus, if $R_{ij} = 0$, then $R_{ik} = 0$ ($j < k$). Let n_i denote the number of observed measurements of participant i .

2.2 Models and Estimation

Let $E(y_{ij}|\mathbf{X}_i) = \mu_{ij}(\boldsymbol{\beta}_0)$ denote the outcome model with an arbitrary link function (e.g., logit link), where $\boldsymbol{\beta}_0$ represents the u -dimensional coefficients of the mean regression of y_{ij} on \mathbf{X}_i . Our main interest is to estimate the parameter of the treatment effect at the last visit of the longitudinal study ($j=T$). The proposed method allows multiple missingness and imputation models, derives a weight including the multiple model information based on the EL method, and then solves the empirical likelihood weighted estimating equations for $\boldsymbol{\beta}_0$. Let $\boldsymbol{\beta}_{EL}$ denote the proposed estimator for $\boldsymbol{\beta}_0$. If any one of the multiple models is correct, $\boldsymbol{\beta}_0$ can be consistently estimated. In this study, S missingness models and K imputation models are considered, which correspond to

modeling $P(R_{ij} = 1 | \mathbf{X}_i, \bar{\mathbf{Z}}_{i,(j-1)}, \boldsymbol{\alpha}^s)$ for $s = 1, \dots, S$ and $P(y_{ij} = 1 | \mathbf{X}_i, \bar{\mathbf{Z}}_{i,(j-1)}, \boldsymbol{\gamma}^k)$ for $k = 1, \dots, K$, respectively, where $\boldsymbol{\alpha}^s$ and $\boldsymbol{\gamma}^k$ are the corresponding parameters. Note that the imputation models can utilize the post-randomization data (i.e., auxiliary variables $\bar{\mathbf{Z}}_{i,(j-1)}$) and are specified at each time point; thus, they differ from the outcome model $\mu_{ij}(\boldsymbol{\beta}_0)$. In the proposed method, two types of (w)GEE for $\boldsymbol{\beta}_0$ were performed, other than $\boldsymbol{\beta}_{EL}$, to derive an EL weight. Let $\boldsymbol{\beta}_{wGEE}^s$ and $\boldsymbol{\beta}_{GEE}^k$ denote the parameters of the (w)GEE models. The parameters $\boldsymbol{\beta}_{wGEE}^s$, $\boldsymbol{\beta}_{GEE}^k$, and $\boldsymbol{\beta}_{EL}$ are identical for the correct missingness or imputation model. A detailed explanation of each model can be found in Sections 2.3 and 2.4.

2.3 Missingness Model $\pi_{ij}^s(\boldsymbol{\alpha}^s)$ and wGEE Model $\mu_{ij}(\boldsymbol{\beta}_{wGEE}^s)$

For the missingness model, let $v_{ij}^s(\boldsymbol{\alpha}^s)$ denote the conditional probability of observing y_{ij} being conditional on $R_{i,(j-1)} = 1$, \mathbf{X}_i and $\bar{\mathbf{Z}}_{i,(j-1)}$, $v_{ij}^s(\boldsymbol{\alpha}^s) = P(R_{ij} = 1 | R_{i,(j-1)} = 1, \mathbf{X}_i, \bar{\mathbf{Z}}_{i,(j-1)}, \boldsymbol{\alpha}^s)$ for $s = 1, \dots, S$, and a set of multiple missingness models as $\text{logit}\{v_{ij}^s(\boldsymbol{\alpha}^s)\} = (\boldsymbol{\alpha}_1^s)^T \mathbf{X}_i + (\boldsymbol{\alpha}_2^s)^T \bar{\mathbf{Z}}_{i,(j-1)}$. We define $P(R_{ij} = 1 | \mathbf{X}_i, \bar{\mathbf{Z}}_{i,(j-1)}, \boldsymbol{\alpha}^s) = \pi_{ij}^s(\boldsymbol{\alpha}^s) = v_{i1}^s(\boldsymbol{\alpha}^s) \times \dots \times v_{ij}^s(\boldsymbol{\alpha}^s)$ as the observation probability for $s = 1, \dots, S$ (Fitzmaurice et al. 1995; Robins et al. 1995).

Using $\pi_{ij}^s(\hat{\boldsymbol{\alpha}}^s)$, the wGEE analysis (Robins et al. 1995) is performed. The estimating equations for $\boldsymbol{\beta}_{wGEE}^s$ for $s = 1, \dots, S$ are as follows:

$$\mathbf{S}^S(\hat{\boldsymbol{\alpha}}^S, \boldsymbol{\beta}_{\text{WGEE}}^S) = \sum_i \frac{\partial \{\boldsymbol{\mu}_i(\boldsymbol{\beta}_{\text{WGEE}}^S)\}^T}{\partial \boldsymbol{\beta}_{\text{WGEE}}^S} (\mathbf{V}_i)^{-1} \mathbf{W}_{\text{WGEE},i}^S \{\mathbf{Y}_i - \boldsymbol{\mu}_i(\boldsymbol{\beta}_{\text{WGEE}}^S)\} = \mathbf{0}, \quad (1)$$

where $\boldsymbol{\mu}_i(\boldsymbol{\beta}_{\text{WGEE}}^S) = \{\mu_{i1}(\boldsymbol{\beta}_{\text{WGEE}}^S), \dots, \mu_{iT}(\boldsymbol{\beta}_{\text{WGEE}}^S)\}^T$, $\mathbf{W}_{\text{WGEE},i}^S = \text{diag}(R_{ij}/\pi_{ij}^S(\hat{\boldsymbol{\alpha}}^S))_{j \leq T}$ as a diagonal matrix factor, \mathbf{V}_i is the working covariance matrix of \mathbf{Y}_i which is assumed as $\mathbf{V}_i = \mathbf{A}_i^{\frac{1}{2}} \mathbf{R}(\boldsymbol{\rho}) \mathbf{A}_i^{\frac{1}{2}}$, where \mathbf{A}_i is a $T \times T$ diagonal matrix whose j th diagonal element is the variance functions of $\mu_{ij}(\hat{\boldsymbol{\beta}}_{\text{WGEE}}^S)$, and $\mathbf{R}(\boldsymbol{\rho})$ is a $T \times T$ working correlation matrix with $\rho_{jj} = 1$ and $\rho_{jj'}$ as a correlation between time points j and j' . The predicted value $\mu_{i,(j-1)}(\hat{\boldsymbol{\beta}}_{\text{WGEE}}^S)$ is used to derive a weight \hat{w}_{ij} (see Section 2.5). When the missingness model includes $y_{i,(j-1)}$ as $\bar{\mathbf{Z}}_{i,(j-1)}$, a missing $y_{i,(j-1)}$ is imputed with the predicted value $\mu_{i,(j-1)}(\hat{\boldsymbol{\beta}}_{\text{WGEE}}^S)$.

2.4 Imputation Model $\phi_{ijd}^k(\boldsymbol{\gamma}_{jd}^k)$ and GEE Model $\boldsymbol{\mu}_{ij}(\boldsymbol{\beta}_{\text{GEE}}^k)$

For the imputation model, to derive the expected value of y_{ij} for each participant and time point, we use the sequential imputation method of Paik (1997), in which it is assumed that

$$E(y_{ij} | \mathbf{X}_i, \bar{\mathbf{Z}}_{i,(j-1)}, R_{i,(j-1)} = 1, R_{ij} = 0) = E(y_{ij} | \mathbf{X}_i, \bar{\mathbf{Z}}_{i,(j-1)}, R_{i,(j-1)} = 1, R_{ij} = 1). \quad (2)$$

Following equation (2), under the MAR assumption, a missing y_{ij} can be imputed using a regression model for $E(y_{ij} | \mathbf{X}_i, \bar{\mathbf{Z}}_{i,(j-1)}, R_{i,(j-1)} = 1, R_{ij} = 1)$. This model is fitted to

y_{ij} , which is either observed or has already been imputed. Because a monotone missing data pattern is assumed, the maximum number of missing patterns is T . As shown in Figure 1, a $T \times T$ table is constructed, where the rows are the missing patterns and the columns are the time points j . Each y_{ij} is assigned to each cell, and the shaded cells are missing y_{ij} . Sequential imputation is performed as in the following procedure.

- (1) As per equation (2), the missing data in the first diagonal cells (1,2), (2,3), ..., (T-1,T) are estimated using the observed data (i.e., $E(y_{ij} | \mathbf{X}_i, \bar{\mathbf{Z}}_{i,(j-1)}, R_{i,(j-1)} = 1, R_{ij} = 1)$). For the imputation of cell (2,3), for example, the data at time point 3, i.e., cells (3,3), (4,3), ..., (T, 3), are used in the regression model. In the first step, cell (1,3) is not imputed.
- (2) The missing data in the second diagonal cells (1,3), (2,4), ..., (T-2,T) cannot be estimated directly because $E(y_{ij} | \mathbf{X}_i, \bar{\mathbf{Z}}_{i,(j-2)}, R_{i,(j-2)} = 1, R_{i,(j-1)} = 0)$ is not equal to $E(y_{ij} | \mathbf{X}_i, \bar{\mathbf{Z}}_{i,(j-1)}, R_{i,(j-1)} = 1, R_{ij} = 1)$. If the imputed values of the first diagonal are treated as observed ones, the missing data in the second diagonal cells (1,3), (2,4), ..., (T-2,T) are estimated using the observed data and the imputed values of the first diagonal in accordance with equation (2). For the imputation of cell (1,3), for example, the data at time point 3, i.e., cells (2,3), (3,3), ..., (T, 3), are used in the regression model.

(3) By imputing y_{ij} with the observed or imputed data, this procedure is repeated along the diagonals until all missing measurements have been imputed.

Let $\phi_{ijd}^k(\mathbf{Y}_{jd}^k)$ ($k = 1, \dots, K$) denote a set of multiple outcome regression models for estimating a predicted value of y_{ij} ,

$$\phi_{ijd}^k(\mathbf{Y}_{jd}^k) = P(y_{ij} = 1 | \mathbf{X}_i, \bar{\mathbf{Z}}_{i,(j-1)}, \mathbf{Y}_{jd}^k), \quad (3)$$

where $d (= 1, \dots, D)$ is the notation for a diagonal pattern.

Using $\phi_{ijd}^k(\hat{\mathbf{Y}}_{jd}^k)$, the GEE analysis is also performed. The estimating equations for $\boldsymbol{\beta}_{\text{GEE}}^k$ for $k = 1, \dots, K$ are as follows:

$$\mathbf{S}^k(\boldsymbol{\beta}_{\text{GEE}}^k, \hat{\mathbf{Y}}_{jd}^k) = \sum_i \frac{\partial \{\boldsymbol{\mu}_i(\boldsymbol{\beta}_{\text{GEE}}^k)\}^T}{\partial \boldsymbol{\beta}_{\text{GEE}}^k} (\mathbf{V}_i)^{-1} \{\hat{\mathbf{Y}}_i - \boldsymbol{\mu}_i(\boldsymbol{\beta}_{\text{GEE}}^k)\} = \mathbf{0}, \quad (4)$$

where $\boldsymbol{\mu}_i(\boldsymbol{\beta}_{\text{GEE}}^k) = \{\mu_{i1}(\boldsymbol{\beta}_{\text{GEE}}^k), \dots, \mu_{iT}(\boldsymbol{\beta}_{\text{GEE}}^k)\}^T$, $\hat{\mathbf{Y}}_i = (\hat{y}_{i1}, \dots, \hat{y}_{iT})^T$. The function

$\sum_i \frac{\partial \{\boldsymbol{\mu}_i(\boldsymbol{\beta}_{\text{GEE}}^k)\}^T}{\partial \boldsymbol{\beta}_{\text{GEE}}^k} (\mathbf{V}_i)^{-1} \{\hat{\mathbf{Y}}_i - \boldsymbol{\mu}_i(\boldsymbol{\beta}_{\text{GEE}}^k)\}$ can be decomposed into

$\sum_i \{\mathbf{S}_{i1}^k(\boldsymbol{\beta}_{\text{GEE}}^k, \hat{\mathbf{Y}}_{jd}^k) + \dots + \mathbf{S}_{iT}^k(\boldsymbol{\beta}_{\text{GEE}}^k, \hat{\mathbf{Y}}_{jd}^k)\}$, where $\mathbf{S}_{ij}^k(\boldsymbol{\beta}_{\text{GEE}}^k, \hat{\mathbf{Y}}_{jd}^k)$ is an element of the

estimating function for each time point. A missing y_{ij} is imputed using $\phi_{ijd}^k(\hat{\mathbf{Y}}_{jd}^k)$

derived from equation (3); thus, the response variable is $\hat{y}_{ij} = R_{ij}y_{ij} + (1 -$

$R_{ij})\phi_{ijd}^k(\hat{\mathbf{Y}}_{jd}^k)$. Note that $\hat{\boldsymbol{\beta}}_{\text{GEE}}^k$ is estimated as if the entire data were available, except

for the fact that missing y_{ij} are replaced by $\phi_{ijd}^k(\hat{\mathbf{Y}}_{jd}^k)$. The estimate $\hat{\boldsymbol{\beta}}_{\text{GEE}}^k$ differs from

$\widehat{\boldsymbol{\beta}}_{\text{WGEE}}^s$, which is estimated using information on missingness model $\pi_{ij}^s(\boldsymbol{\alpha}^s)$. Thus,

$\widehat{\boldsymbol{\beta}}_{\text{WGEE}}^s$ and $\widehat{\boldsymbol{\beta}}_{\text{GEE}}^k$ are derived from different informations (i.e., missingness models and imputation models). After $\widehat{\boldsymbol{\beta}}_{\text{GEE}}^k$ is estimated, the function $\mathbf{S}_i^k(\widehat{\boldsymbol{\beta}}_{\text{GEE}}^k, \widehat{\boldsymbol{\gamma}}_{jd}^k) =$

$$\frac{\partial \{\boldsymbol{\mu}_i(\widehat{\boldsymbol{\beta}}_{\text{GEE}}^k)\}^T}{\partial \boldsymbol{\beta}_{\text{GEE}}^k} (\mathbf{V}_i)^{-1} \{\widehat{\mathbf{Y}}_i - \boldsymbol{\mu}_i(\widehat{\boldsymbol{\beta}}_{\text{GEE}}^k)\} = \mathbf{S}_{i1}^k(\widehat{\boldsymbol{\beta}}_{\text{GEE}}^k, \widehat{\boldsymbol{\gamma}}_{1d}^k) + \dots + \mathbf{S}_{iT}^k(\widehat{\boldsymbol{\beta}}_{\text{GEE}}^k, \widehat{\boldsymbol{\gamma}}_{Td}^k)$$
 is

calculated to derive a weight \widehat{w}_{ij} (see Section 2.5). The response variables are replaced by the expected value as $\widehat{\mathbf{Y}}_i = \{\widehat{y}_{i1}, \phi_{i2d}^k(\widehat{\boldsymbol{\gamma}}_{2d}^k), \dots, \phi_{iTd}^k(\widehat{\boldsymbol{\gamma}}_{Td}^k)\}^T$. An observed value y_{i1} is replaced with \widehat{y}_{i1} , which is the estimate of the expected value of y_{i1} conditioned on \mathbf{X}_i . In addition, each observed value y_{ij} (for $j=2, \dots, T$) is replaced with $\phi_{ij1}^k(\widehat{\boldsymbol{\gamma}}_{j1}^k)$, which is derived from the imputation model described by equation (3) for the first diagonal cell.

2.5 Derivation of Weight \widehat{w}_{ij} Using EL Method

The proposed method assigns a weight w_{ij} to each observation at all time points (using the element-wise EL method) (Wang et al. 2010). To derive a weight \widehat{w}_{ij} for the proposed estimator, the following constraints are imposed:

$$\sum_{i=1}^N \sum_{j=1}^{n_i} w_{ij} = 1, w_{ij} \geq 0,$$

$$\sum_{i=1}^N \sum_{j=1}^{n_i} w_{ij} \widehat{\mathbf{g}}_{ij}(\widehat{\boldsymbol{\alpha}}, \widehat{\boldsymbol{\beta}}_{\text{GEE}}, \widehat{\boldsymbol{\beta}}_{\text{WGEE}}, \widehat{\boldsymbol{\gamma}}_{jd}) = \mathbf{0}, \quad (5)$$

where $\hat{\mathbf{g}}_{ij}(\hat{\boldsymbol{\alpha}}, \hat{\boldsymbol{\beta}}_{\text{GEE}}, \hat{\boldsymbol{\beta}}_{\text{wGEE}}, \hat{\boldsymbol{\gamma}}_{jd}) =$

$$[\pi_{ij}^1(\hat{\boldsymbol{\alpha}}^1) - \theta^1(\hat{\boldsymbol{\alpha}}^1, \hat{\boldsymbol{\beta}}_{\text{wGEE}}^1), \dots, \pi_{ij}^S(\hat{\boldsymbol{\alpha}}^S) - \theta^S(\hat{\boldsymbol{\alpha}}^S, \hat{\boldsymbol{\beta}}_{\text{wGEE}}^S),$$

$$\{\mathbf{s}_{ij}^1(\hat{\boldsymbol{\beta}}_{\text{GEE}}^1, \hat{\boldsymbol{\gamma}}_{jd}^1) - \boldsymbol{\xi}^1(\hat{\boldsymbol{\beta}}_{\text{GEE}}^1, \hat{\boldsymbol{\gamma}}_{jd}^1)\}^T, \dots, \{\mathbf{s}_{ij}^K(\hat{\boldsymbol{\beta}}_{\text{GEE}}^K, \hat{\boldsymbol{\gamma}}_{jd}^K) - \boldsymbol{\xi}^K(\hat{\boldsymbol{\beta}}_{\text{GEE}}^K, \hat{\boldsymbol{\gamma}}_{jd}^K)\}^T]^T,$$

$$\theta^S(\hat{\boldsymbol{\alpha}}^S, \hat{\boldsymbol{\beta}}_{\text{wGEE}}^S) = \sum_{i=1}^N \sum_{j=1}^T \pi_{ij}^S(\hat{\boldsymbol{\alpha}}^S, \hat{\boldsymbol{\beta}}_{\text{wGEE}}^S) / (NT), \text{ and}$$

$$\boldsymbol{\xi}^k(\hat{\boldsymbol{\beta}}_{\text{GEE}}^k, \hat{\boldsymbol{\gamma}}_{jd}^k) = \sum_{i=1}^N \sum_{j=1}^T \mathbf{s}_{ij}^k(\hat{\boldsymbol{\beta}}_{\text{GEE}}^k, \hat{\boldsymbol{\gamma}}_{jd}^k) / (NT).$$

The equation can be also described as follows:

$$\sum_{i=1}^N \sum_{j=1}^{n_i} w_{ij} \{\pi_{ij}^S(\hat{\boldsymbol{\alpha}}^S) - \theta^S(\hat{\boldsymbol{\alpha}}^S, \hat{\boldsymbol{\beta}}_{\text{wGEE}}^S)\} = 0 \quad (s = 1, \dots, S), \text{ and}$$

$$\sum_{i=1}^N \sum_{j=1}^{n_i} w_{ij} \{\mathbf{s}_{ij}^k(\hat{\boldsymbol{\beta}}_{\text{GEE}}^k, \hat{\boldsymbol{\gamma}}_{jd}^k) - \boldsymbol{\xi}^k(\hat{\boldsymbol{\beta}}_{\text{GEE}}^k, \hat{\boldsymbol{\gamma}}_{jd}^k)\} = \mathbf{0} \quad (k = 1, \dots, K).$$

When $j \geq n_i + 1$, w_{ij} is defined as zero. Vector $\hat{\mathbf{g}}_{ij}$ includes S functions for missingness model $\pi_{ij}^S(\hat{\boldsymbol{\alpha}}^S)$ and K functions for GEE $\mathbf{s}_{ij}^k(\hat{\boldsymbol{\beta}}_{\text{GEE}}^k, \hat{\boldsymbol{\gamma}}_{jd}^k)$; i.e., the missing data are imputed using a set of K imputation models. Similar to Han (2014), our proposed method employs the calibration method (Lumley et al. 2011). Calibration weights are defined by matching the values of the calibration variables based on the sampled participants to the corresponding known population values. In this setting, $\pi_{ij}^S(\hat{\boldsymbol{\alpha}}^S)$ and $\mathbf{s}_{ij}^k(\hat{\boldsymbol{\beta}}_{\text{GEE}}^k, \hat{\boldsymbol{\gamma}}_{jd}^k)$ play the role of calibration variables and their population values $\theta^S(\hat{\boldsymbol{\alpha}}^S, \hat{\boldsymbol{\beta}}_{\text{wGEE}}^S)$ and $\boldsymbol{\xi}^k(\hat{\boldsymbol{\beta}}_{\text{GEE}}^k, \hat{\boldsymbol{\gamma}}_{jd}^k)$ are estimated by the unweighted sample

averages over all the participants at all the time points ($j = 1, \dots, T$). The weight $\{\widehat{w}_{ij}: i = 1, \dots, N, j = 1, \dots, n_i\}$ is a set of weights assigned to the biased sample $\{y_{ij}: i = 1, \dots, N, j = 1, \dots, n_i\}$. Through equation (5), w_{ij} corrects for the selection bias so that certain population quantities may be consistently estimated based on the biased sample. When a missingness model includes $y_{i,(j-1)}$ as $\mathbf{Z}_{i,(j-1)}$, the missing $y_{i,(j-1)}$ are imputed using $\mu_{i,(j-1)}(\widehat{\boldsymbol{\beta}}_{\text{wGEE}}^s)$ derived from equation (1) $\{\widehat{y}_{i,(j-1)} = R_{i,(j-1)}y_{i,(j-1)} + (1 - R_{i,(j-1)})\mu_{i,(j-1)}(\widehat{\boldsymbol{\beta}}_{\text{wGEE}}^s)\}$. Thus, the predicted value $\pi_{ij}^s(\widehat{\boldsymbol{\alpha}}^s, \widehat{\boldsymbol{\beta}}_{\text{wGEE}}^s)$ is estimated using $\mu_{i,(j-1)}(\widehat{\boldsymbol{\beta}}_{\text{wGEE}}^s)$.

Let $m = \sum_{i=1}^N \sum_{j=1}^T R_{ij}$ denote the total number of observed measurements over the time points. By maximizing $\prod_{i=1}^N \prod_{j=1}^{n_i} w_{ij}$ subject to the constraints in equation (5) using Lagrange multipliers, the following set of equations can be obtained:

$$\widehat{w}_{ij} = \frac{1}{m} \times \frac{1}{1 + \widehat{\boldsymbol{\lambda}}^T \widehat{\boldsymbol{g}}_{ij}(\widehat{\boldsymbol{\alpha}}, \widehat{\boldsymbol{\beta}}_{\text{GEE}}, \widehat{\boldsymbol{\beta}}_{\text{wGEE}}, \widehat{\boldsymbol{\gamma}}_{jd})}, \quad (6)$$

$$\sum_{i=1}^N \sum_{j=1}^{n_i} \widehat{w}_{ij} \widehat{\boldsymbol{g}}_{ij}(\widehat{\boldsymbol{\alpha}}, \widehat{\boldsymbol{\beta}}_{\text{GEE}}, \widehat{\boldsymbol{\beta}}_{\text{wGEE}}, \widehat{\boldsymbol{\gamma}}_{jd}) = \mathbf{0},$$

$$\frac{1}{m} \sum_{i=1}^N \sum_{j=1}^{n_i} \frac{\widehat{\boldsymbol{g}}_{ij}(\widehat{\boldsymbol{\alpha}}, \widehat{\boldsymbol{\beta}}_{\text{GEE}}, \widehat{\boldsymbol{\beta}}_{\text{wGEE}}, \widehat{\boldsymbol{\gamma}}_{jd})}{1 + \widehat{\boldsymbol{\lambda}}^T \widehat{\boldsymbol{g}}_{ij}(\widehat{\boldsymbol{\alpha}}, \widehat{\boldsymbol{\beta}}_{\text{GEE}}, \widehat{\boldsymbol{\beta}}_{\text{wGEE}}, \widehat{\boldsymbol{\gamma}}_{jd})} = \mathbf{0}, \quad (7)$$

$$\widehat{\boldsymbol{\lambda}} = \arg \min_{\boldsymbol{\lambda} \in \mathcal{D}_m} \mathbf{F}_m(\boldsymbol{\lambda}),$$

where $\boldsymbol{\lambda}$ is an $(S + uK)$ -dimensional Lagrange multiplier vector, corresponding to the

functions $\hat{\mathbf{g}}_{ij}(\hat{\boldsymbol{\alpha}}, \hat{\boldsymbol{\beta}}_{\text{GEE}}, \hat{\boldsymbol{\beta}}_{\text{wGEE}}, \hat{\boldsymbol{\gamma}}_{jd})$ (Owen 2001). To estimate $\boldsymbol{\lambda}$, we define $\mathbf{F}_m(\boldsymbol{\lambda}) = -m^{-1} \sum_{i=1}^N \sum_{j=1}^{n_i} \log\{1 + \boldsymbol{\lambda}^T \hat{\mathbf{g}}_{ij}(\hat{\boldsymbol{\alpha}}, \hat{\boldsymbol{\beta}}_{\text{GEE}}, \hat{\boldsymbol{\beta}}_{\text{wGEE}}, \hat{\boldsymbol{\gamma}}_{jd})\}$ and let $\mathcal{D}_m = \{\boldsymbol{\lambda}: 1 + \boldsymbol{\lambda}^T \hat{\mathbf{g}}_{ij}(\hat{\boldsymbol{\alpha}}, \hat{\boldsymbol{\beta}}_{\text{GEE}}, \hat{\boldsymbol{\beta}}_{\text{wGEE}}, \hat{\boldsymbol{\gamma}}_{jd}) > 0\}$ (Han 2014). It is easy to see that $\partial \mathbf{F}_m(\boldsymbol{\lambda}) / \partial \boldsymbol{\lambda} = \mathbf{0}$ because of equation (7). The estimate $\hat{\boldsymbol{\lambda}}$ minimizes $\mathbf{F}_m(\boldsymbol{\lambda})$ in \mathcal{D}_m . Here, the existence and uniqueness of $\hat{\boldsymbol{\lambda}}$ are guaranteed by the strict convexity of $\mathbf{F}_m(\boldsymbol{\lambda})$ on \mathcal{D}_m . Let $\pi_{ij}^1(\boldsymbol{\alpha}^1)$ denote a correctly specified missingness model and $\boldsymbol{\alpha}_0^1$ denote the true value of $\boldsymbol{\alpha}^1$. As in the method of Han (2014), $\hat{\boldsymbol{\lambda}} = (\hat{\lambda}_1, \dots, \hat{\lambda}_{S+uK})^T$ can be rewritten as $\hat{\lambda}_1 = \frac{\hat{\kappa}_1 + 1}{\theta^1(\hat{\boldsymbol{\alpha}}^1, \hat{\boldsymbol{\beta}}_{\text{wGEE}}^1)}$, $\hat{\lambda}_l = \frac{\hat{\kappa}_l}{\theta^1(\hat{\boldsymbol{\alpha}}^1, \hat{\boldsymbol{\beta}}_{\text{wGEE}}^1)}$ $\{l = 2, \dots, (S + uK)\}$, where $\boldsymbol{\kappa}$ is an $(S + uK)$ -dimensional Lagrange multiplier vector defined in Appendix A and $\hat{\boldsymbol{\kappa}}$ converges to $\mathbf{0}$ in probability when any one of the missingness models is correctly specified. Thus, $\hat{\boldsymbol{w}}_{ij}$ can be also described as follows:

$$\hat{\boldsymbol{w}}_{ij} = \frac{1}{m} \times \frac{1}{1 + \hat{\boldsymbol{\lambda}}^T \hat{\mathbf{g}}_{ij}(\hat{\boldsymbol{\alpha}}, \hat{\boldsymbol{\beta}}_{\text{GEE}}, \hat{\boldsymbol{\beta}}_{\text{wGEE}}, \hat{\boldsymbol{\gamma}}_{jd})} = \frac{1}{m} \times \frac{\theta^1(\hat{\boldsymbol{\alpha}}^1, \hat{\boldsymbol{\beta}}_{\text{wGEE}}^1) / \pi_{ij}^1(\hat{\boldsymbol{\alpha}}^1)}{1 + \hat{\boldsymbol{\kappa}}^T \hat{\mathbf{g}}_{ij}(\hat{\boldsymbol{\alpha}}, \hat{\boldsymbol{\beta}}_{\text{GEE}}, \hat{\boldsymbol{\beta}}_{\text{wGEE}}, \hat{\boldsymbol{\gamma}}_{jd}) / \pi_{ij}^1(\hat{\boldsymbol{\alpha}}^1)}.$$

The details are provided in Appendix A.

2.6 Proposed Estimator: EL wGEE

Using the weights $\hat{\boldsymbol{w}}_{ij}$ described in the previous section, we define the EL wGEE as

$$\mathbf{S}(\boldsymbol{\beta}_{\text{EL}}) = \sum_{i=1}^N \frac{\partial \{\boldsymbol{\mu}_i(\boldsymbol{\beta}_{\text{EL}})\}^T}{\partial \boldsymbol{\beta}_{\text{EL}}} (\mathbf{V}_i)^{-1} \mathbf{W}_{\text{EL},i} \{\mathbf{Y}_i - \boldsymbol{\mu}_i(\boldsymbol{\beta}_{\text{EL}})\} = \mathbf{0}. \quad (8)$$

As in equation (1), $\mathbf{W}_{EL,i}$ is a $T \times T$ diagonal matrix with the j th diagonal element being $R_{ij}\widehat{w}_{ij}$. The solution to equation (8) can be obtained using the Newton–Raphson method. As in Han (2014), the proposed estimator has multiple robustness property.

Theorem 1. The proposed estimator $\widehat{\boldsymbol{\beta}}_{EL}$ is consistent if any one of a set of S missingness models or K imputation models is correctly specified (see Appendixes A and B).

As several models need to be defined and estimates need to be prepared to execute the proposed method, the procedure is summarized as follows:

- (1) Define S missingness models $\pi_{ij}^s(\widehat{\boldsymbol{\alpha}}^s)$ and K imputation models $\phi_{ijd}^k(\widehat{\boldsymbol{\gamma}}_{jd}^k)$.
- (2) If auxiliary variables are included in the missingness model, perform the wGEE analysis with $\pi_{ij}^s(\widehat{\boldsymbol{\alpha}}^s)$, and calculate $\mu_{i,(j-1)}(\widehat{\boldsymbol{\beta}}_{wGEE}^s)$.
- (3) Perform GEE analysis where the missing response variable y_{ij} is imputed with $\phi_{ijd}^k(\widehat{\boldsymbol{\gamma}}_{jd}^k)$, and calculate $\mathbf{S}_{ij}^k(\widehat{\boldsymbol{\beta}}_{GEE}^k, \widehat{\boldsymbol{\gamma}}_{jd}^k)$.
- (4) Using $\pi_{ij}^s(\widehat{\boldsymbol{\alpha}}^s, \widehat{\boldsymbol{\beta}}_{wGEE}^s)$ and $\mathbf{S}_{ij}^k(\widehat{\boldsymbol{\beta}}_{GEE}^k, \widehat{\boldsymbol{\gamma}}_{jd}^k)$ for all the participants at all the time points, calculate $\theta^s(\widehat{\boldsymbol{\alpha}}^s, \widehat{\boldsymbol{\beta}}_{wGEE}^s)$, $\boldsymbol{\xi}^k(\widehat{\boldsymbol{\beta}}_{GEE}^k, \widehat{\boldsymbol{\gamma}}_{jd}^k)$, and $\widehat{\boldsymbol{g}}_{ij}(\widehat{\boldsymbol{\alpha}}, \widehat{\boldsymbol{\beta}}_{GEE}, \widehat{\boldsymbol{\beta}}_{wGEE}, \widehat{\boldsymbol{\gamma}}_{jd})$ using equation (5).
- (5) Estimate $\widehat{\boldsymbol{\lambda}}$ using equation (7), and calculate weight \widehat{w}_{ij} using equation (6).
- (6) Perform an EL wGEE analysis with weight \widehat{w}_{ij} , and obtain $\widehat{\boldsymbol{\beta}}_{EL}$.

If no missingness model is specified, $\pi_{ij}^s(\hat{\boldsymbol{\alpha}}^s)$, $\mu_{i,(j-1)}(\hat{\boldsymbol{\beta}}_{\text{wGEE}}^s)$, and $\theta^s(\hat{\boldsymbol{\alpha}}^s, \hat{\boldsymbol{\beta}}_{\text{wGEE}}^s)$ are not calculated. In addition, if no imputation model is specified, $\phi_{ijd}^k(\hat{\boldsymbol{\gamma}}_{jd}^k)$, $\mathbf{s}_{ij}^k(\hat{\boldsymbol{\beta}}_{\text{GEE}}^k, \hat{\boldsymbol{\gamma}}_{jd}^k)$, and $\boldsymbol{\xi}^k(\hat{\boldsymbol{\beta}}_{\text{GEE}}^k, \hat{\boldsymbol{\gamma}}_{jd}^k)$ are not calculated.

To derive the asymptotic variance of $\hat{\boldsymbol{\beta}}_{\text{EL}}$, as in the method of Han (2014), the first-order Taylor expansion of the EL wGEE around $(\mathbf{0}^T, \boldsymbol{\alpha}_*^T, \boldsymbol{\beta}_{\text{GEE},*}^T, \boldsymbol{\gamma}_{jd,*}^T, \boldsymbol{\beta}_0^T)$ is a possible approach (Tsiatis 2006), where $\boldsymbol{\alpha}_*^T = \{(\boldsymbol{\alpha}_*^1)^T, \dots, (\boldsymbol{\alpha}_*^S)^T\}$, $\boldsymbol{\beta}_{\text{GEE},*}^T = \{(\boldsymbol{\beta}_{\text{GEE},*}^1)^T, \dots, (\boldsymbol{\beta}_{\text{GEE},*}^K)^T\}$, and $\boldsymbol{\gamma}_{jd,*}^T = \{(\boldsymbol{\gamma}_{jd,*}^1)^T, \dots, (\boldsymbol{\gamma}_{jd,*}^K)^T\}$ denote the probability limits of $\hat{\boldsymbol{\alpha}}^T = \{(\hat{\boldsymbol{\alpha}}^1)^T, \dots, (\hat{\boldsymbol{\alpha}}^S)^T\}$, $\hat{\boldsymbol{\beta}}_{\text{GEE}}^T = \{(\hat{\boldsymbol{\beta}}_{\text{GEE}}^1)^T, \dots, (\hat{\boldsymbol{\beta}}_{\text{GEE}}^K)^T\}$, and $\hat{\boldsymbol{\gamma}}_{jd}^T = \{(\hat{\boldsymbol{\gamma}}_{jd}^1)^T, \dots, (\hat{\boldsymbol{\gamma}}_{jd}^K)^T\}$, and the parameter $\boldsymbol{\beta}_0$ is defined as the true value. We derived two formulas for the asymptotic variance of $\hat{\boldsymbol{\beta}}_{\text{EL}}$ under the assumption that $\boldsymbol{\kappa}$, $\boldsymbol{\alpha}$, and $\boldsymbol{\gamma}_{jd}$ are known/unknown parameters. The detailed explanation is presented in Web Appendix A of Supporting information. Under the assumption that each parameter is an unknown parameter and all auxiliary variables are observed, we derived the asymptotic variance estimator \mathbf{V}_{emp1} and confirmed that the comparison of efficiency between $\hat{\boldsymbol{\beta}}_{\text{EL}}$ and $\hat{\boldsymbol{\beta}}_{\text{wGEE}}$ is inconclusive (see Web Appendix A of Supporting information). As the asymptotic variance estimator \mathbf{V}_{emp1} is valid only when any one of the missingness models is correctly specified, we additionally derived the asymptotic variance estimator \mathbf{V}_{emp2} under the assumption that $\boldsymbol{\kappa}$, $\boldsymbol{\alpha}$, and $\boldsymbol{\gamma}_{jd}$ are known parameters.

Theorem 2. Under the assumption that $\boldsymbol{\kappa}$, $\boldsymbol{\alpha}$, and $\boldsymbol{\gamma}_{jd}$ are known parameters,

$$\sqrt{N}(\widehat{\boldsymbol{\beta}}_{\text{EL}} - \boldsymbol{\beta}_0) \xrightarrow{D} N \left[0, \left\{ E \left(\frac{\partial \boldsymbol{Q}_i}{\partial \boldsymbol{\beta}} \right) \right\}^{-1} E(\boldsymbol{Q}_i \boldsymbol{Q}_i^T) \left\{ \left(E \left(\frac{\partial \boldsymbol{Q}_i}{\partial \boldsymbol{\beta}} \right) \right)^{-1} \right\}^T \right]$$

$$= N \left[0, E(\boldsymbol{B}_i)^{-1} E(\boldsymbol{Q}_i \boldsymbol{Q}_i^T) \{E(\boldsymbol{B}_i)^{-1}\}^T \right], \quad (9)$$

where $\boldsymbol{Q}_i = \frac{\partial \{\boldsymbol{\mu}_i(\widehat{\boldsymbol{\beta}}_{\text{EL}})\}^T}{\partial \boldsymbol{\beta}} (\boldsymbol{V}_i)^{-1} \boldsymbol{W}_{\text{EL},i} \{\boldsymbol{Y}_i - \boldsymbol{\mu}_i(\widehat{\boldsymbol{\beta}}_{\text{EL}})\}$ and

$$\boldsymbol{B}_i = \frac{\partial \{\boldsymbol{\mu}_i(\widehat{\boldsymbol{\beta}}_{\text{EL}})\}^T}{\partial \boldsymbol{\beta}} (\boldsymbol{V}_i)^{-1} \boldsymbol{W}_{\text{EL},i} \frac{\partial \boldsymbol{\mu}_i(\widehat{\boldsymbol{\beta}}_{\text{EL}})}{\partial \boldsymbol{\beta}}.$$

We also suggest a bootstrapping method as alternative to calculate the variance of $\widehat{\boldsymbol{\beta}}_{\text{EL}}$.

3 Simulation Studies

We performed Monte Carlo simulations to evaluate the performance of the proposed method and estimated the bias, standard error (SE), mean square error (MSE), and type-1 error rate. We set the sample size n to 75 or 150 in each treatment group with 1000 Monte Carlo simulations. In the simulation study, we primarily investigated the following considerations: (1) whether EL wGEE exhibits multiple robustness compared to the bias of logistic regression and GEE; (2) whether the efficiency of EL wGEE is comparable to that of wGEE when the missingness model is correctly specified; and (3) whether EL wGEE exhibits adequate performance compared to that of AIPW when

multiple models are specified.

3.1 Simulation Datasets

We generated the full dataset for correlated binary outcomes y_{ij} ($j = 1, 2, 3, 4$) using Lee's method (1997). The marginal model for y_{ij} was defined as $\text{logit}\{P(y_{ij} = 1)\} = \beta_0 + \beta_{\text{TRT}}X_{\text{TRT},i} + \beta_{1j} + \beta_{2j}X_{\text{TRT},i} + \beta_3X_{\text{BL},i}$. The model had four independent design variables: the treatment $X_{\text{TRT},i}$ (group A: 1, group B: 0), time points, interactions between the treatment and time points, and baseline covariate $X_{\text{BL},i}$, which was generated from $N(0, 2)$. The parameters $\beta_{\text{TRT}} = 1.5$, $\beta_{14} = \beta_{24} = 0$, $\beta_3 = 0.2$, and the other parameters were obtained from the given marginal response rates over time, and were considered under two scenarios as follows:

Scenario	Response rate at time points 1, 2, 3, 4			
Low-response	0.25	0.31	0.38	0.48 for group A
	0.14	0.15	0.17	0.17 for group B
High-response	0.27	0.43	0.73	0.82 for group A
	0.18	0.21	0.50	0.50 for group B

The within-subject correlation structure was set as the autoregressive (1) type, and the correlation parameter ρ was set to 0.5, which was set based on the plaque psoriasis study data. In this simulation, our primary objective was to evaluate the

estimation efficiency (bias, SE, and MSE) of the treatment effect at $j = 4$, which corresponds to $\beta_{\text{TRT}} = 1.5$. In addition, a type-1 error rate was evaluated by additionally generating null-hypothesis datasets for $\beta_{\text{TRT}} = 0$, $\beta_{21} = \beta_{22} = \beta_{23} = \beta_{24} = 0$.

We assumed a monotone missing data pattern, and the true missingness model was $\text{logit}\{P(R_{ij} = 1 | R_{i,(j-1)} = 1)\} = \alpha_0 + \alpha_1 X_{\text{TRT},i} + \alpha_2 y_{i,(j-1)} + \alpha_3 z_{i,(j-1)}$, where R_{ij} denotes the indicator variable for observing y_{ij} and z_{ij} , and $y_{i,(j-1)}$ and $z_{i,(j-1)}$ are auxiliary variables. An auxiliary variable z_{ij} ($j = 1, 2, 3$) was generated from $y_{ij} + \varepsilon_{ij}$, $\varepsilon_{ij} \sim N(0, 0.5)$. The indicator variables R_{ij} were generated from six different missingness models determined by $\alpha = (\alpha_0, \alpha_1, \alpha_2, \alpha_3)$. The six missing data mechanisms were given by (pt1) $\alpha = (\alpha_0, -1, -1, -1)$, (pt2) $\alpha = (\alpha_0, -1, 1, 1)$, (pt3) $\alpha = (\alpha_0, 1, -1, -1)$, (pt4) $\alpha = (\alpha_0, 0, -1, -1)$, (pt5) $\alpha = (\alpha_0, 1, 1, 1)$, and (pt6) $\alpha = (\alpha_0, 0, 1, 1)$, where the magnitudes of parameters α_2 and α_3 were set based on the plaque psoriasis study's auxiliary variables. In addition, we considered two missing data proportions at the last time point $j = 4$ as follows: (i) 15% (moderate missing data scenario) and (ii) 30% (high missing data scenario). These two missing data scenarios were adjusted from α_0 .

3.2 Estimators

We specified a correct/misspecified model for the missingness model and the

imputation model for y_{ij} , respectively. The correct missingness model $\pi_{ij}^1(\boldsymbol{\alpha}^1)$ was defined as $\text{logit}\{P(R_{ij} = 1 | R_{i,(j-1)} = 1)\} = \alpha_0^1 + \alpha_1^1 X_{\text{TRT},i} + \alpha_2^1 y_{i,(j-1)} + \alpha_3^1 z_{i,(j-1)}$, where missing $y_{i,(j-1)}$ and $z_{i,(j-1)}$ were imputed using predicted values derived from wGEE in the proposed method (Section 2.5), and the correct imputation model $\phi_{ij,d}^1(\boldsymbol{\gamma}_{jd}^1)$ was defined as $\text{logit}\{P(y_{ij} = 1)\} = \gamma_{0,jd}^1 + \gamma_{1,jd}^1 X_{\text{TRT},i} + \gamma_{2,jd}^1 X_{\text{BL},i} + (\boldsymbol{\gamma}_{3,jd}^1)^T \bar{\mathbf{Y}}_{i,(j-1)}$, where $\bar{\mathbf{Y}}_{i,(j-1)} = (\hat{y}_{i1}, \dots, \hat{y}_{i,(j-1)})^T$, $\hat{y}_{i,(j-1)} = R_{i,(j-1)} y_{i,(j-1)} + (1 - R_{i,(j-1)}) \phi_{i,(j-1),d}^1(\hat{\boldsymbol{\gamma}}_{(j-1),d}^1)$. To achieve asymptotic unbiasedness, it is important to include auxiliary variable $\bar{\mathbf{Y}}_{i,(j-1)}$ in the imputation model because, under MAR, $\bar{\mathbf{Y}}_{i,(j-1)}$ contains information on the missing data y_{ij} . The misspecified missingness model $\pi_{ij}^2(\boldsymbol{\alpha}^2)$ was defined as $\text{logit}\{P(R_{ij} = 1 | R_{i,(j-1)} = 1)\} = \alpha_0^2 + \alpha_1^2 X_{1,i}$, and the misspecified imputation model $\phi_{ij,d}^2(\boldsymbol{\gamma}_{jd}^2)$ was defined as $\text{logit}\{P(y_{ij} = 1)\} = \gamma_{0,jd}^2 + \gamma_{1,jd}^2 X_{\text{TRT},i} + \gamma_{2,jd}^2 X_{1,i}$. The covariate $X_{1,i}$ was generated as $\exp(\zeta_1)$, where $\zeta_1 \sim N(0.1, 0.5)$ for the misspecified model. In the GEEs model, we fit the correctly specified model as $\text{logit}\{P(y_{ij} = 1)\} = \beta_0 + \beta_{\text{TRT}} X_{\text{TRT},i} + \beta_{1j} + \beta_{2j} X_{\text{TRT},i} + \beta_3 X_{\text{BL},i}$, except that we specified the correlation structure as unstructured and imposed the constraint that the parameter for time point 4 be equal to zero ($\beta_{14} = \beta_{24} = 0$) in order to make $\hat{\beta}_{\text{TRT}}$ the treatment effect at time point 4.

For comparison with the proposed estimator (EL wGEE), β_{TRT} was also estimated using a logistic regression model (CLUDE), GEE (Liang and Zeger 1986),

wGEE (Robins et al. 1995), and the AIPW method (Seaman and Copas 2009). In CLUDE, the outcome model was defined as $\text{logit}\{\mu_{i4}(\mathbf{X}_i, \boldsymbol{\beta})\} = \beta_{\text{INT}} + \beta_{\text{TRT}}X_{\text{TRT},i} + \beta_{\text{BL}}X_{\text{BL},i}$.

3.3 Simulation Results

For each of the 48 ways that data were generated (Low/High-response rate \times six missing data mechanisms \times 15%/30% missing data proportion \times two pattern of sample size for 75/150 per treatment group), we compared the five methods of analysis. Of these six missing data mechanisms, all analysis methods using the coarsened datasets from (pt5) $\alpha = (\alpha_0, 1, 1, 1)$ and (pt6) $\alpha = (\alpha_0, 0, 1, 1)$ had unbiased estimates of β_{TRT} , thus, the results for these two scenarios are not shown.

When estimating the parameters of missingness/imputation models, quasi-complete separation occurred in some simulation datasets. Those datasets were therefore excluded from the analysis. The number of quasi-complete separations is shown in Web Appendix B of Supporting information. We consider that the impact of excluding these datasets on the performance evaluation of each analysis method is small. We checked the distribution of the remaining datasets (i.e., this excluded the datasets in which quasi-complete separation occurred) and found that the mean values of $\hat{\beta}_{\text{TRT}}$ in the complete datasets were similar among the 1000 Monte Carlo datasets and

the remaining datasets (the difference in the mean values of $\hat{\beta}_{\text{TRT}}$ ranged from 0.0001 to 0.001). This implies that there are no characteristic differences among the 1000 Monte Carlo datasets and the remaining datasets. In addition, the quasi-complete separations in each scenario (except missing mechanism of pattern 2) were approximately less than 10 out of 1000 datasets; thus, we consider that the performance of each estimator could be evaluated with the remaining datasets. We report the simulation results for the scenario of 30% missing data proportion and $n = 150$ because the pattern of bias, SE, MSE, and type-1 error rate in the scenario of 15% missing data proportion and $n = 75$ were similar, but the bias was smaller in the scenario of 15% missing data proportion. These results for the scenario of 15% missing data proportion and $n = 75$ are summarized in the Web Appendix. As discussed by Emerson and Owen (2009) and Han (2014), the empirical likelihood method may be problematic when the sample size is small, in which case $\mathbf{0}$ may not be in the convex hull of $\hat{\mathbf{g}}_{ij}(\hat{\boldsymbol{\alpha}}, \hat{\boldsymbol{\beta}}_{\text{GEE}}, \hat{\boldsymbol{\beta}}_{\text{wGEE}}, \hat{\boldsymbol{\gamma}}_{jd})$. In the simulation studies, there was no difficulty with estimating parameter $\hat{\boldsymbol{\lambda}}$.

Figure 2 reports on the bias of $\hat{\beta}_{\text{TRT}}$ for $n = 150$ and 30% missing data proportion. The notation of these estimators includes the names of the estimators and the model specification, which have the form “method-0000.” Each digit in the four-digit number, from left to right, indicates whether or not

$\pi_{ij}^1(\boldsymbol{\alpha}^1)$, $\pi_{ij}^2(\boldsymbol{\alpha}^2)$, $\phi_{ijd}^1(\boldsymbol{\gamma}_{jd}^1)$, or $\phi_{ijd}^2(\boldsymbol{\gamma}_{jd}^2)$ is included, where “1” indicates “included” and “0” indicates “not included.” In the simulation results, EL wGEE exhibited better

performance when a correct model was included among the missingness or imputation models compared to when only misspecified models were included, excluding pt3 of the low-response rate datasets.

wGEE exhibited similar performance to that of EL wGEE under the correctly specified missingness model.

AIPW exhibited similar performance to that of EL wGEE in most of the scenarios, especially when one missingness model and one imputation model are specified and both models are correctly specified (i.e., EL wGEE-1010 and AIPW-1010). The similarity of the estimators is explained in Appendix C. Interestingly, the bias for AIPW was larger than that for EL wGEE when only the imputation model was correctly specified (i.e., AIPW-0110) in the pt1, pt3, and pt4 of the high-response rate datasets. In addition, the bias for AIPW was larger than that for EL wGEE when both missingness and imputation models were misspecified (AIPW-0101). When there is a higher risk of misspecification for the missingness model, EL wGEE may be more robust than AIPW and wGEE.

Although GEE resulted in biased estimates in most of the scenarios, as expected, GEE interestingly exhibited better performance than EL wGEE in pt3 of the low-response rate datasets and pt2 of the high-response rate datasets. The pt2 ($\alpha = (\alpha_0, -1, 1, 1)$) describes a situation in which subjects who are treated with an active drug are vulnerable to dropout because of an adverse event and an outcome is assumed to be

an indicator of improvement (i.e., if $y_{ij} = 0$, the subject is at a higher risk of dropout). The pt3 ($\alpha = (\alpha_0, 1, -1, -1)$) describes a situation in which subjects who are treated with a placebo is vulnerable to dropout because of lack of efficacy and an outcome is assumed as an indicator of worsening of symptoms (i.e., if $y_{ij} = 1$, the subject is at a higher risk of dropout). In addition, GEE also exhibited better performance than wGEE when the missingness model was misspecified. Thus, in some scenarios, GEE might be applied in missing longitudinal data analysis.

Web Tables 5 to 8 of Supporting information report on the SEs of $\hat{\beta}_{\text{TRT}}$ ($\sum_{i=1}^{1000} SE_{\hat{\beta}_{\text{TRT},i}}/1000$). The SEs of EL wGEE were calculated based on equation (9). The SEs of AIPW were also calculated using the sandwich estimator under the assumption that both the missingness and imputation models are correct (Seaman and Copas 2009, p. 944). In addition, the empirical SE, which is the standard deviation in a 1000 replication parameter estimate $\hat{\beta}_{\text{TRT}}$, is shown to confirm the validity of these SEs. Although the SEs of EL wGEE, AIPW, and wGEE were similar to or slightly larger than those of CLUDE and GEE, the difference became smaller in the scenario of 15% missing data proportion (Web Tables 6 and 8). We expected that the efficiency (SEs) of the EL wGEE method would be improved owing to use of the longitudinal response variable; however, adding weights to the estimating equations may have a greater influence on the estimation efficiency.

Figure 3 reports on the MSEs of $\hat{\beta}_{\text{TRT}}$ for $n = 150$ and 30% missing data proportion $\left(\sum_{i=1}^{1000}(\hat{\beta}_{\text{TRT},i} - \beta_{\text{TRT}})^2 / 1000\right)$. Overall, although the results of MSEs were similar to that of SEs, the MSEs of EL wGEE were comparative to those of CLUDE even under 30% missing data proportion. Other results are summarized in Web Tables 9 to 12 of Supporting information. The results indicate that the degree of inefficiency of EL wGEE relative to the CLUDE or GEE estimator depends on the proportion of missing data. At least, the impact on efficiency was small in applying the proposed methods in this setting.

Figure 4 reports on the Type-1 error rate (using null-hypothesis datasets for $\beta_{\text{TRT}} = 0, \beta_{21} = \beta_{22} = \beta_{23} = \beta_{24} = 0$). In general, the Type-I error rates of EL wGEE approximately ranged from 4% to 6% and were similar to those of other methods in each scenario.

4 Application: Plaque Psoriasis Phase III Study

In this section, we describe the application of the proposed method to the Plaque Psoriasis Phase III Study. In total, 542 participants were randomized in a 2:1 ratio to the M518101 or placebo group. Efficacy assessments were taken at baseline, Weeks 2, 4, 6, and 8. The primary endpoint was the success rate based on the IGA score at Week 8. In other efficacy outcomes, we evaluated the modified Psoriasis Area and Severity Index

(mPASI, in the range 0-no disease to 64.8-maximal disease) and itch score (on a scale from 0-None to 3-Severe). The treatment effect $\hat{\beta}_{\text{TRT}}$ was estimated using the CLUDE for the complete case, GEE, wGEE, AIPW and the proposed estimator (EL wGEE). Let X_{TRT} denote the treatment (M518101 = 1, placebo = 0) and X_{BL} the baseline mPASI score. We used three auxiliary variables $\mathbf{Z} = (Z_1, Z_2, Z_3)$, where Z_1, Z_2 , and Z_3 indicate the success rate according to the IGA score, mPASI score, and itch score at a previous visit, respectively.

It is desirable to include all the components of the auxiliary variables and the interactions in one missingness model and one imputation model; however, this was not feasible in this case, due to the quasi-complete separation for the binary outcome model. Nevertheless, the proposed estimator was able to address this issue via multiple modeling. The missingness model was $\text{logit}\{P(R_{ij} = 1 | R_{i,(j-1)} = 1, \mathbf{X}_i, \mathbf{Z}_{i,(j-1)})\} = \alpha_0^s + \alpha_1^s X_{\text{TRT},i} + \alpha_2^s X_{\text{BL},i} + \alpha_3^s + \alpha_4^s X_{\text{TRT},i} \times X_{\text{BL},i} + \alpha_5^s \mathbf{Z}_{i,(j-1)}$, where the two missingness models ($s = 1, 2$) were set as follows: $\pi_{ij}^1(\boldsymbol{\alpha}^1)$ with auxiliary variables Z_1, Z_2 , and $X_{\text{TRT}} \times Z_2$, and $\pi_{ij}^2(\boldsymbol{\alpha}^2)$ with an auxiliary variables Z_3 , and $X_{\text{TRT}} \times Z_3$. The imputation model for y_{ij} was $\text{logit}\{P(y_{ij} = 1 | \mathbf{X}_i, \mathbf{Z}_{i,(j-1)})\} = \phi_{ijd}^k(\boldsymbol{\gamma}_{jd}^k) = \boldsymbol{\gamma}_{0jd}^k + \boldsymbol{\gamma}_{1jd}^k X_{\text{TRT},i} + \boldsymbol{\gamma}_{2jd}^k X_{\text{BL},i} + \boldsymbol{\gamma}_{3jd}^k \mathbf{Z}_{i,(j-1)}$. As with the missingness model, the two imputation models ($k = 1, 2$) were set as follows: $\phi_{ijd}^1(\boldsymbol{\gamma}_{jd}^1)$ with auxiliary variables

Z_1 , and Z_2 , and $\phi_{ijd}^2(\mathbf{Y}_{jd}^2)$ with an auxiliary variable Z_3 . The GEE analysis was performed using the same settings as in Section 3.

As running example, the proposed method was performed as follows:

- (1) Estimated the parameter of missingness models for $v_{ij}^1(\boldsymbol{\alpha}^1)$ and calculated values of $\pi_{ij}^1(\hat{\boldsymbol{\alpha}}^1)$ in each participant and time point $\{i = 1, \dots, N, j = 1, \dots, n_i\}$.
The same analysis was performed with $v_{ij}^2(\boldsymbol{\alpha}^2)$, and calculated values of $\pi_{ij}^2(\hat{\boldsymbol{\alpha}}^2)$.
- (2) Performed wGEE analysis with a combination of weight $R_{ij}/\pi_{ij}^1(\hat{\boldsymbol{\alpha}}^1)$ and outcome \mathbf{Z}_1 , where the outcome model was defined as $\text{logit}\{P(Z_{1,ij} = 1)\} = \beta_{\text{WGEE},0}^1 + \beta_{\text{WGEE},\text{TRT}}^1 X_{\text{TRT},i} + \beta_{\text{WGEE},1j}^1 + \beta_{\text{WGEE},2j}^1 X_{\text{TRT},i} + \beta_{\text{WGEE},3}^1 X_{\text{BL},i}$.
- (3) Calculated predicted values of \mathbf{Z}_1 using the wGEE model.
- (4) The procedures described in (2) and (3) performed again with a combination of $R_{ij}/\pi_{ij}^1(\hat{\boldsymbol{\alpha}}^1)$ and \mathbf{Z}_2 as well as $R_{ij}/\pi_{ij}^2(\hat{\boldsymbol{\alpha}}^2)$ and \mathbf{Z}_3 to calculate predicted values of \mathbf{Z}_2 and \mathbf{Z}_3 .
- (5) After missing \mathbf{Z}_1 , \mathbf{Z}_2 , and \mathbf{Z}_3 were imputed with the predicted values of (3) and (4), $\pi_{ij}^s(\hat{\boldsymbol{\alpha}}^s, \hat{\boldsymbol{\beta}}_{\text{WGEE}}^s)$, $s = 1, 2$ for all the participants at all the time points $\{i = 1, \dots, N, j = 1, \dots, T\}$ were calculated. These estimates $\pi_{ij}^s(\hat{\boldsymbol{\alpha}}^s, \hat{\boldsymbol{\beta}}_{\text{WGEE}}^s)$, $s = 1, 2$ were used in (10).

- (6) Performed Paik's imputation method using the imputation model $\phi_{ijd}^1(\mathbf{Y}_{jd}^1)$ and calculated predicted values $\phi_{ijd}^1(\hat{\mathbf{Y}}_{jd}^1)$ of y_{ij} in each participant and time point. The same Paik's imputation was performed using the imputation model $\phi_{ijd}^2(\mathbf{Y}_{jd}^2)$. Similar to $\phi_{ijd}^k(\mathbf{Y}_{jd}^k)$, missing \mathbf{Z}_2 were imputed with the predicted values of $Z_{2,ij}$ using the imputation model for $Z_{2,ij}$ with auxiliary variables $y_{i,(j-1)}$ and $Z_{2,i,(j-1)}$; missing \mathbf{Z}_3 were imputed with the predicted values of $Z_{3,ij}$ using the imputation model for $Z_{3,ij}$ with auxiliary variables $y_{i,(j-1)}$ and $Z_{3,i,(j-1)}$.
- (7) Performed the GEE analysis, where the outcome model was defined as $\text{logit}\{P(y_{ij} = 1)\} = \beta_{\text{GEE},0}^1 + \beta_{\text{GEE},\text{TRT}}^1 X_{\text{TRT},i} + \beta_{\text{GEE},1j}^1 + \beta_{\text{GEE},2j}^1 X_{\text{TRT},i} + \beta_{\text{GEE},3}^1 X_{\text{BL},i}$, and the missing response variable y_{ij} was imputed with $\phi_{ijd}^1(\hat{\mathbf{Y}}_{jd}^1)$.
- (8) Calculated the function $\mathbf{S}_{ij}^1(\hat{\boldsymbol{\beta}}_{\text{GEE}}^1, \hat{\mathbf{Y}}_{jd}^1)$, where all values y_{ij} , including the observed ones, were replaced by model-based expectation $\phi_{ijd}^1(\hat{\mathbf{Y}}_{jd}^1)$. The calculated $\mathbf{S}_{ij}^1(\hat{\boldsymbol{\beta}}_{\text{GEE}}^1, \hat{\mathbf{Y}}_{jd}^1)$ were used in (10).
- (9) The procedures described in (7) and (8) performed again with $\phi_{ijd}^2(\hat{\mathbf{Y}}_{jd}^2)$ and $\hat{\boldsymbol{\beta}}_{\text{GEE}}^2$, and finally calculated $\mathbf{S}_{ij}^2(\hat{\boldsymbol{\beta}}_{\text{GEE}}^2, \hat{\mathbf{Y}}_{jd}^2)$.

- (10) Using $\pi_{ij}^s(\hat{\alpha}^s, \hat{\beta}_{\text{wGEE}}^s)$ and $S_{ij}^k(\hat{\beta}_{\text{GEE}}^k, \hat{\gamma}_{jd}^k)$, $s = 1, 2, k = 1, 2$ for all the participants at all the time points, calculated $\theta^s(\hat{\alpha}^s, \hat{\beta}_{\text{wGEE}}^s)$, $\xi^k(\hat{\beta}_{\text{GEE}}^k, \hat{\gamma}_{jd}^k)$, $s = 1, 2, k = 1, 2$ and $\hat{g}_{ij}(\hat{\alpha}, \hat{\beta}_{\text{GEE}}, \hat{\beta}_{\text{wGEE}}, \hat{\gamma}_{jd})$ using equation (5).
- (11) Estimated $\hat{\lambda}$ using equation (7), and calculated weight \hat{w}_{ij} using equation (6).
- (12) Performed an EL wGEE analysis with weight \hat{w}_{ij} , and obtained $\hat{\beta}_{\text{EL}}$.

GEE analysis (Liang and Zeger 1986) and wGEE analysis (Robins et al. 1995) correspond to the EL wGEE analysis in step (12), where the weight \hat{w}_{ij} is replaced by identity matrix, $R_{ij}/\pi_{ij}^1(\hat{\alpha}^1)$ and $R_{ij}/\pi_{ij}^2(\hat{\alpha}^2)$, respectively.

The results are summarized in Table 1. In Table 1, the odds ratio of CLUDE was 4.62, while that of EL wGEE was 3.83 to 4.08, and that of AIPW was 3.93 to 4.05, respectively. We speculated that these estimators mitigated the bias of CLUDE. In addition, these findings were more deeply evaluated by the multiple modeling of EL wGEE-1111, which included all components of auxiliary variables. Therefore, the proposed method was useful for evaluating the stability of the conclusion.

5 Discussion and Conclusions

The proposed method extends Han's method to the longitudinal outcome model. The simulation results indicate that the EL wGEE method achieves better performance if any one of the specified missingness or imputation models is correct (multiple

robustness).

In these simulation studies, we also evaluated the performance of this method when multiple missingness and imputation models were utilized. Greenland and Pearce (2015, p.99) state that “Methods that model both outcome and exposure (including doubly robust methods) avoid having to make a choice, but at the cost of more modeling effort. They have the option of using more data information with potential accuracy gains as a result.” Although diagnostic measures have been widely used in model assessment, it is difficult to apply them to incomplete outcome and auxiliary data; consequently, there is significant concern regarding model misspecification in such a situation. In addition, when a dimension of auxiliary variables is high (i.e., there are many auxiliary variables), there is a risk of (quasi-) complete separation for the binary outcome model. In such cases, multiple lower dimensional models can be used to resolve the issues. However, there have been only a few studies on the performance evaluation of the proposed method when misspecified models are included; thus, it might not be preferable to use multiple models without a strong reason. We recommend specifying one missingness model and/or one imputation model for the primary analysis and additional models for the sensitivity analysis. The need for sensitivity analysis regarding missing data is established, and its importance is emphasized in the ICH E9 (R1) (2019). Notably, the multiple modeling method allows for a wide range of sensitivity analyses.

Although the imputation approach in the proposed method was under MAR assumption, the approach can be modified to be under missing not at random (MNAR), such as “jump-to-reference (J2R)” to perform a sensitivity analysis. In J2R (Mallinckrodt et al., 2017), missing data for reference group are imputed assuming MAR; missing data for drug-treated group are imputed assuming MNAR such that the benefit from the drug immediately disappears after discontinuation of the study drug. Therefore, parameter γ_{jd} was estimated using data only for the reference group. Note that since this analysis is under MNAR, specifying missingness models $\pi_{ij}^s(\hat{\alpha}^s)$ under MAR is not reasonable.

In some simulation datasets, especially those with low-response rate and $n=75$, quasi-complete separation occurred. In the scenario of pt2 ($\alpha = (\alpha_0, -1, 1, 1)$) and low-response rate, the phenomenon was often observed because most subjects drop out after $y_{ij} = 0$. In this situation, where it is difficult to estimate the parameters of the missingness/imputation models, we may choose logistic regression or the GEE method—for which there is no need to specify missingness/imputation models.

In Web Appendix C, we additionally compare the performance of EL wGEE to that of Han’s method, in which we specify an imputation and missingness model similar to that of the proposed method and use the imputation method of Paik. If explanatory variables $y_{i,(j-1)}$ and $z_{i,(j-1)}$ were missing in the missingness model of Han’s method, we imputed them using last observation carried forward (LOCF) method to calculate

$\theta^1 (= \sum_{i=1}^N \pi_{i4}^1(\hat{\alpha}^1)/N)$. The EL wGEE method achieved better performance with Bias, as compared to that of Han's method with LOCF in such an eventuality. It appears that the coarsened explanatory variables $\bar{\mathbf{Z}}_{ij}$ (i. e., y_{i3} and z_{i3}) affect the accuracy of $\hat{\beta}_{\text{TRT}}$. When a subject i drops out at or before time point 3, the $\pi_{i4}^1(\hat{\alpha}^1)$ cannot be estimated, because the explanatory variable $\bar{\mathbf{Z}}_{i3}$ (i. e., y_{i3} and z_{i3}) is missing. In addition, the observation probability $\pi_{i4}^1(\hat{\alpha}^1)$ is estimated for all of the participants to calculate $\theta^1 (= \sum_{i=1}^N \pi_{i4}^1(\hat{\alpha}^1)/N)$. To cope with this problem, $\pi_{i4}^1(\hat{\alpha}^1)$ was estimated substituting y_{i2} or y_{i1} in place of y_{i3} in Han's method with LOCF in our simulation studies (the same applies to z_{i3}), but that might introduce bias. In contrast, the proposed method can impute the missing $\bar{\mathbf{Z}}_{i3}$ (i. e., y_{i3} and z_{i3}) with the predicted values derived from wGEE. The proposed method can therefore be used for longitudinal data analysis.

In the psoriasis clinical trial data analysis, although it was not feasible to include all the components of the auxiliary variables as well as the interactions caused by the quasi-complete separation of the binary outcome model, such as in wGEE and AIPW. The proposed estimator was able to address this issue via multiple modeling and enable the further evaluation of the validity of the conclusions derived.

Although we assumed that the outcome is binary data and that \mathbf{Z}_{ij} includes only y_{ij} to simplify the explanation presented in Section 2, the proposed method can easily be extended to continuous and count data outcomes and used to handle multiple auxiliary variables. In terms of multiple robustness, the proposed method is expected to

be more robust and useful than existing methods when the dimension of the auxiliary variables is high, and/or reasonable multiple candidate models are available. Future work should further investigate the relative performance of the other methods under a wider range of settings.

Acknowledgments

We thank Maruho Co., Ltd., Japan, for providing the plaque psoriasis phase III trial data. We are deeply grateful to Dr. Katsuhiko Omae for invaluable advice provided.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This work was supported by the Japan Agency for Medical Research and Development [grant number JP16lk0201061].

Conflict of interest

Hiroshi Komazaki is an employee of Maruho Co., Ltd.

Masaaki Doi is an employee of Ono Pharmaceutical Co., Ltd.

Naohiro Yonemoto is an employee of Pfizer Japan Inc.

Supporting information

Additional supporting information for the web appendices and tables referenced in Sections 2.6, 3.3, and 5 may be found online in Supporting Information section at the end of the article.

References

- Bang, H., and Robins, J. M. (2005), “Doubly robust estimation in missing data and causal inference models.” *Biometrics*, 61, 962–973. <https://doi.org/10.1111/j.1541-0420.2005.00377.x>.
- Emerson, S. C., and Owen, A. B. (2009), “Calibration of the empirical likelihood method for a vector mean.” *Electronic Journal of Statistics*, 3, 1161–1192. <https://doi.org/10.1214/09-EJS518>.
- Fitzmaurice, G. M., Molenberghs, G., and Lipsitz S. R. (1995), “Regression models for longitudinal binary responses with informative drop-outs.” *Journal of the Royal Statistical Society: Series B (Methodological)*, 57, 691–704. <https://doi.org/10.1111/j.2517-6161.1995.tb02056.x>.
- Greenland, S., and Pearce, N. (2015), “Statistical foundations for model-based adjustments.” *Annual Review of Public Health*, 36, 89–108. <https://doi.org/10.1146/annurev-publhealth-031914-122559>.
- Han, P., and Wang, L. (2013), “Estimation with missing data: Beyond double robustness.” *Biometrika*, 100, 417–430. <https://doi.org/10.1093/biomet/ass087>.
- Han, P. (2014), “Multiply robust estimation in regression analysis with missing data.” *Journal of the American Statistical Association*, 109, 1159–1173. <https://doi.org/10.1080/01621459.2014.880058>.

Han, P., Song, P. X. K., and Wang, L. (2015), “Achieving semiparametric efficiency bound in longitudinal data analysis with dropouts.” *Journal of Multivariate Analysis*, 135, 59–70. <https://doi.org/10.1016/j.jmva.2014.12.006>.

Han, P. (2016), “Intrinsic efficiency and multiple robustness in longitudinal studies with drop-out.” *Biometrika*, 103, 683–700. <https://doi.org/10.1093/biomet/asw024>.

Horvitz, D. G., and Thompson, D. J. (1952), “A generalization of sampling without replacement from a finite universe.” *Journal of the American Statistical Association*, 47, 663–685. <https://doi.org/10.1080/01621459.1952.10483446>.

ICH Guideline E9 (R1): Addendum on Estimands and sensitivity analysis in clinical trials. Final version. Published 20 November 2019.

https://database.ich.org/sites/default/files/E9-R1_Step4_Guideline_2019_1203.pdf
(accessed 23 October 2022).

Kang, J. D., and Schafer, J. L. (2007), “Demystifying double robustness: A comparison of alternative strategies for estimating a population mean from incomplete data.” *Statistical Science*, 22, 523–539. <https://doi.org/10.1214/07-STS227>.

Lee, A. J. (1997), “Some simple methods for generating correlated categorical variates.” *Computational Statistics & Data Analysis*, 26, 133–148. [https://doi.org/10.1016/S0167-9473\(97\)00030-3](https://doi.org/10.1016/S0167-9473(97)00030-3).

Liang, K. Y., and Zeger, S. L. (1986), “Longitudinal data analysis using generalized linear models.” *Biometrika*, 73, 13–22. <https://doi.org/10.1093/biomet/73.1.13>.

Little, R. J., and Rubin, D. B. (2002), "Statistical analysis with missing data." John Wiley and Sons.

Lumley, T., Shaw, P. A., and Dai, J. Y. (2011), "Connections between survey calibration estimators and semiparametric models for incomplete data." *International Statistical Review*, 79, 200–220. <https://doi.org/10.1111/j.1751-5823.2011.00138.x>.

Mallinckrodt, C. H. (2013), "*Preventing and treating missing data in longitudinal clinical trials: A practical guide*." Cambridge University Press.

Mallinckrodt, C., Molenberghs, G., and Rathmann, S. (2017), "Choosing estimands in clinical trials with missing data." *Pharmaceutical Statistics*, 16, 29–36.

<https://doi.org/10.1002/pst.1765>.

Owen, A. B. (2001), "*Empirical likelihood*." Chapman and Hall/CRC Press.

Paik, M. C. (1997), "The generalized estimating equation approach when data are not missing completely at random." *Journal of the American Statistical Association*, 92, 1320–1329. <https://doi.org/10.1080/01621459.1997.10473653>.

Robins, J. M., Rotnitzky, A., and Zhao, L. P. (1995), "Analysis of semiparametric regression models for repeated outcomes in the presence of missing data." *Journal of the American Statistical Association*, 90, 106–121.

<https://doi.org/10.1080/01621459.1995.10476493>.

Seaman, S., and Copas, A. (2009), “Doubly robust generalized estimating equations for longitudinal data.” *Statistics in Medicine*, 28, 937–955.

<https://doi.org/10.1002/sim.3520>.

Tsiatis, A. A. (2006), “*Semiparametric theory and missing data*.” New York: Springer.

Wang, S., Qian, L., and Carroll, R. J. (2010), “Generalized empirical likelihood methods for analyzing longitudinal data.” *Biometrika*, 97, 79–93.

<https://doi.org/10.1093/biomet/asp073>.

Appendix

Appendix A: Connection Between $\hat{\lambda}$ and $\hat{\kappa}$

To prove that $\hat{\beta}_{EL}$ is consistent if any one of the missingness models is correctly specified in Appendix B, we provide the derivation of the weight \hat{w}_{ij} subject to different constraints.

More formally, we denote

$$\mathbf{g}_j(\boldsymbol{\alpha}_*, \boldsymbol{\beta}_{GEE,*}, \boldsymbol{\gamma}_{jd,*}) = [\pi_j^1(\boldsymbol{\alpha}_*^1) - \theta_*^1(\boldsymbol{\alpha}_*^1), \dots, \pi_j^S(\boldsymbol{\alpha}_*^S) - \theta_*^S(\boldsymbol{\alpha}_*^S), \\ \{\mathbf{s}_j^1(\boldsymbol{\beta}_{GEE,*}^1, \boldsymbol{\gamma}_{jd,*}^1) - \boldsymbol{\xi}^1(\boldsymbol{\beta}_{GEE,*}^1, \boldsymbol{\gamma}_{jd,*}^1)\}^T, \dots, \{\mathbf{s}_j^K(\boldsymbol{\beta}_{GEE,*}^K, \boldsymbol{\gamma}_{jd,*}^K) - \boldsymbol{\xi}^K(\boldsymbol{\beta}_{GEE,*}^K, \boldsymbol{\gamma}_{jd,*}^K)\}^T]^T.$$

The expected value $E\{\sum_{j=1}^T R_{ij} \mathbf{g}_j(\boldsymbol{\alpha}_*, \boldsymbol{\beta}_{GEE,*}, \boldsymbol{\gamma}_{jd,*})\} \neq \mathbf{0}$, when the missing data mechanism is MAR; thus, the Lagrange multipliers $\boldsymbol{\lambda}$ are also not guaranteed to converge to **zero** based on the theory of empirical likelihood (Owen 2001). Let $\pi_{ij}^1(\boldsymbol{\alpha}^1)$ denote a correctly specified missingness model, $\boldsymbol{\alpha}_0^1$ denote the true value of $\boldsymbol{\alpha}^1$, and p_{ij} denote the conditional empirical observation probability on $R_{ij} = 1$, where p_{ij} is subjected to the following constraints:

$$\sum_{i=1}^N \sum_{j=1}^{n_i} p_{ij} = 1,$$

$$p_{ij} \geq 0,$$

$$\sum_{i=1}^N \sum_{j=1}^{n_i} p_{ij} \{\pi_{ij}^s(\hat{\boldsymbol{\alpha}}^s) - \theta^s(\hat{\boldsymbol{\alpha}}^s, \hat{\boldsymbol{\beta}}_{wGEE}^s)\} / \pi_{ij}^1(\hat{\boldsymbol{\alpha}}^1) = 0 \quad (s = 1, \dots, S), \text{ and}$$

$$\sum_{i=1}^N \sum_{j=1}^{n_i} p_{ij} \{\mathbf{s}_{ij}^k(\hat{\boldsymbol{\beta}}_{GEE}^k, \hat{\boldsymbol{\gamma}}_{jd}^k) - \boldsymbol{\xi}^k(\hat{\boldsymbol{\beta}}_{GEE}^k, \hat{\boldsymbol{\gamma}}_{jd}^k)\} / \pi_{ij}^1(\hat{\boldsymbol{\alpha}}^1) = \mathbf{0} \quad (k = 1, \dots, K).$$

As in equation (6), using the Lagrange multipliers, we obtain

$$\hat{p}_{ij} = \frac{1}{m} \frac{1}{1 + \hat{\kappa}^T \hat{g}_{ij}(\hat{\alpha}, \hat{\beta}_{GEE}, \hat{\beta}_{WGEE}, \hat{\gamma}_{jd}) / \pi_{ij}^1(\hat{\alpha}^1)},$$

where $\hat{\kappa}$ is also an $(S + uK)$ -dimensional Lagrange multiplier vector. When the missingness model is correctly specified, $E\{\sum_{j=1}^T R_{ij} \mathbf{g}_j(\boldsymbol{\alpha}_*, \boldsymbol{\beta}_{GEE,*}, \boldsymbol{\gamma}_{jd,*}) / \pi_{ij}^1(\boldsymbol{\alpha}^1)\} = \mathbf{0}$, and the Lagrange multiplier vector $\hat{\kappa}$ therefore converges to $\mathbf{0}$ (Owen 2001). As in equation (7), the estimating equations for $\boldsymbol{\kappa}$ are as follows:

$$\sum_{i=1}^N \sum_{j=1}^{n_i} p_{ij} \hat{g}_{ij}(\hat{\alpha}, \hat{\beta}_{GEE}, \hat{\beta}_{WGEE}, \hat{\gamma}_{jd}) / \pi_{ij}^1(\hat{\alpha}^1) = \mathbf{0},$$

$$\frac{1}{m} \sum_{i=1}^N \sum_{j=1}^{n_i} \frac{\hat{g}_{ij}(\hat{\alpha}, \hat{\beta}_{GEE}, \hat{\beta}_{WGEE}, \hat{\gamma}_{jd}) / \pi_{ij}^1(\hat{\alpha}^1)}{1 + \hat{\kappa}^T \hat{g}_{ij}(\hat{\alpha}, \hat{\beta}_{GEE}, \hat{\beta}_{WGEE}, \hat{\gamma}_{jd}) / \pi_{ij}^1(\hat{\alpha}^1)} = \mathbf{0}. \quad (\text{A.1})$$

Then,

$$\frac{1}{m} \sum_{i=1}^N \sum_{j=1}^{n_i} \frac{\hat{g}_{ij}(\hat{\alpha}, \hat{\beta}_{GEE}, \hat{\beta}_{WGEE}, \hat{\gamma}_{jd})}{\pi_{ij}^1(\hat{\alpha}^1) + \hat{\kappa}^T \hat{g}_{ij}(\hat{\alpha}, \hat{\beta}_{GEE}, \hat{\beta}_{WGEE}, \hat{\gamma}_{jd})} = \mathbf{0},$$

$$\frac{1}{\theta^1(\hat{\alpha}^1, \hat{\beta}_{WGEE}^1) \times m} \sum_{i=1}^N \sum_{j=1}^{n_i} \frac{\hat{g}_{ij}(\hat{\alpha}, \hat{\beta}_{GEE}, \hat{\beta}_{WGEE}, \hat{\gamma}_{jd})}{1 + \left(\frac{\kappa_1 + 1}{\theta^1(\hat{\alpha}^1, \hat{\beta}_{WGEE}^1)}, \frac{\kappa_2}{\theta^1(\hat{\alpha}^1, \hat{\beta}_{WGEE}^1)}, \dots, \frac{\kappa_{(S+uK)}}{\theta^1(\hat{\alpha}^1, \hat{\beta}_{WGEE}^1)} \right) \hat{g}_{ij}(\hat{\alpha}, \hat{\beta}_{GEE}, \hat{\beta}_{WGEE}, \hat{\gamma}_{jd})} = \mathbf{0}.$$

By comparing the above equation with equation (7), it is evident that $\hat{\lambda}_1 = \frac{\hat{\kappa}_1 + 1}{\theta^1(\hat{\alpha}^1, \hat{\beta}_{WGEE}^1)}$,

$\hat{\lambda}_l = \frac{\hat{\kappa}_l}{\theta^1(\hat{\alpha}^1, \hat{\beta}_{WGEE}^1)}$, $l = 2, \dots, (S+uK)$. The weight \hat{w}_{ij} can be described as follows:

$$\hat{w}_{ij} = \frac{1}{m} \times \frac{1}{1 + \hat{\lambda}^T \hat{g}_{ij}(\hat{\alpha}, \hat{\beta}_{GEE}, \hat{\beta}_{WGEE}, \hat{\gamma}_{jd})} = \frac{1}{m} \times \frac{\theta^1(\hat{\alpha}^1, \hat{\beta}_{WGEE}^1) / \pi_{ij}^1(\hat{\alpha}^1)}{1 + \hat{\kappa}^T \hat{g}_{ij}(\hat{\alpha}, \hat{\beta}_{GEE}, \hat{\beta}_{WGEE}, \hat{\gamma}_{jd}) / \pi_{ij}^1(\hat{\alpha}^1)}. \quad (\text{A.2})$$

Appendix B: Multiple Robustness of the Proposed Estimator

The proposed estimator $\widehat{\boldsymbol{\beta}}_{\text{EL}}$ is consistent if any one of a set of S missingness models or K imputation models is correctly specified. The details are as follows:

When the missingness model is correctly specified

Appendix A demonstrates that the weight \widehat{w}_{ij} can be denoted as shown in equation (A.2).

When $\pi_{ij}^1(\boldsymbol{\alpha}^1)$ is correctly specified, $\widehat{\boldsymbol{\kappa}} \xrightarrow{p} \mathbf{0}$. Therefore, the estimating equations in equation (8) are asymptotically equivalent to the wGEE (Robins et al. 1995), which is weighted with the inverse probability of observing. When the missingness model is correctly specified, the left-hand side of the estimating equations converge to **zero** in probability under the true parameter $\boldsymbol{\beta}_0$. Thus, we have

$$\sum_{i=1}^N \frac{\partial \{\boldsymbol{\mu}_i(\boldsymbol{\beta}_0)\}^T}{\partial \boldsymbol{\beta}} (\mathbf{V}_i)^{-1} \mathbf{W}_{\text{EL},i} \{\mathbf{Y}_i - \boldsymbol{\mu}_i(\boldsymbol{\beta}_0)\} = \mathbf{0},$$

$$\sum_{i=1}^N \sum_{j=1}^{n_i} w_{ij} \mathbf{S}_{ij}(\boldsymbol{\beta}_0) = \mathbf{0},$$

$$\frac{1}{m} \sum_{i=1}^N \sum_{j=1}^{n_i} \frac{\theta^1(\widehat{\boldsymbol{\alpha}}^1, \widehat{\boldsymbol{\beta}}_{\text{wGEE}}^1) / \pi_{ij}^1(\widehat{\boldsymbol{\alpha}}^1)}{1 + \widehat{\boldsymbol{\kappa}}^T \widehat{\boldsymbol{g}}_{ij}(\widehat{\boldsymbol{\alpha}}^1, \widehat{\boldsymbol{\beta}}_{\text{GEE}}, \widehat{\boldsymbol{\beta}}_{\text{wGEE}}, \widehat{\boldsymbol{\gamma}}_{jd}) / \pi_{ij}^1(\widehat{\boldsymbol{\alpha}}^1)} \mathbf{S}_{ij}(\boldsymbol{\beta}_0)$$

$$= \frac{\theta^1(\widehat{\boldsymbol{\alpha}}^1, \widehat{\boldsymbol{\beta}}_{\text{wGEE}}^1)}{m} \sum_{i=1}^N \sum_{j=1}^T \frac{R_{ij}}{\pi_{ij}^1(\widehat{\boldsymbol{\alpha}}^1)} \mathbf{S}_{ij}(\boldsymbol{\beta}_0) + o_p(1)$$

$$\xrightarrow{p} \frac{\theta^1(\boldsymbol{\alpha}_0^1, \boldsymbol{\beta}_{\text{wGEE}}^1)}{\sum_{j=1}^T P(R_{ij}=1)} \mathbb{E} \left\{ \sum_{j=1}^T \frac{R_{ij}}{\pi_{ij}^1(\boldsymbol{\alpha}_0^1)} \mathbf{S}_{ij}(\boldsymbol{\beta}_0) \right\} = \mathbf{0}. \quad (\text{B.1})$$

When the imputation model for y_{ij} is correctly specified

Let $\phi_{ijd}^1(\hat{\boldsymbol{\gamma}}_{jd}^1)$ denote the estimate of the expected value of y_{ij} using a correctly specified imputation model, $\boldsymbol{\gamma}_{jd,0}^1$ denote the true value of $\hat{\boldsymbol{\gamma}}_{jd}^1$, and $\hat{\boldsymbol{\beta}}_{\text{GEE}}^1$ denote the estimated parameter of the GEE model using $\phi_{ijd}^1(\hat{\boldsymbol{\gamma}}_{jd}^1)$. The estimated parameter $\hat{\boldsymbol{\beta}}_{\text{GEE}}^1$ converges to the true parameter $\boldsymbol{\beta}_0$ in probability. From the constraints in equation (5),

$$\sum_{i=1}^N \sum_{j=1}^{n_i} w_{ij} \boldsymbol{S}_{ij}^1(\hat{\boldsymbol{\beta}}_{\text{GEE}}^1, \hat{\boldsymbol{\gamma}}_{jd}^1) = \sum_{i=1}^N \sum_{j=1}^T \boldsymbol{S}_{ij}^1(\hat{\boldsymbol{\beta}}_{\text{GEE}}^1, \hat{\boldsymbol{\gamma}}_{jd}^1)/(NT). \text{ Under this condition,}$$

both sides of the equations converge to **zero** in probability. The main analysis function

$$\sum_{i=1}^N \sum_{j=1}^{n_i} \hat{w}_{ij} \boldsymbol{S}_{ij}(\boldsymbol{\beta}_0) \text{ also therefore converges to } \mathbf{0} \text{ in probability. By using } y_{ij} \perp$$

$$R_{ij} | (\mathbf{X}_{ij}, \bar{\mathbf{Z}}_{i,(j-1)}) \text{ and equation (8),}$$

$$\sum_{i=1}^N \sum_{j=1}^{n_i} \hat{w}_{ij} \boldsymbol{S}_{ij}(\boldsymbol{\beta}_0) = \mathbf{0},$$

$$\sum_{i=1}^N \sum_{j=1}^{n_i} \hat{w}_{ij} \boldsymbol{S}_{ij}(\boldsymbol{\beta}_0) -$$

$$\sum_{i=1}^N \sum_{j=1}^{n_i} \hat{w}_{ij} \boldsymbol{S}_{ij}^1(\hat{\boldsymbol{\beta}}_{\text{GEE}}^1, \hat{\boldsymbol{\gamma}}_{jd}^1) + \sum_{i=1}^N \sum_{j=1}^T \boldsymbol{S}_{ij}^1(\hat{\boldsymbol{\beta}}_{\text{GEE}}^1, \hat{\boldsymbol{\gamma}}_{jd}^1)/(NT) = \mathbf{0},$$

$$\sum_{i=1}^N \sum_{j=1}^{n_i} \hat{w}_{ij} \{ \boldsymbol{S}_{ij}(\boldsymbol{\beta}_0) - \boldsymbol{S}_{ij}^1(\hat{\boldsymbol{\beta}}_{\text{GEE}}^1, \hat{\boldsymbol{\gamma}}_{jd}^1) \} + \sum_{i=1}^N \sum_{j=1}^T \hat{\boldsymbol{S}}_{ij}^1(\hat{\boldsymbol{\beta}}_{\text{GEE}}^1, \hat{\boldsymbol{\gamma}}_{jd}^1)/(NT) = \mathbf{0},$$

$$\frac{1}{m} \sum_{i=1}^N \sum_{j=1}^T \frac{R_{ij} \{ \boldsymbol{S}_{ij}(\boldsymbol{\beta}_0) - \boldsymbol{S}_{ij}^1(\hat{\boldsymbol{\beta}}_{\text{GEE}}^1, \hat{\boldsymbol{\gamma}}_{jd}^1) \}}{1 + \hat{\boldsymbol{\lambda}}^T \hat{\boldsymbol{g}}_{ij}(\hat{\boldsymbol{\alpha}}, \hat{\boldsymbol{\beta}}_{\text{GEE}}, \hat{\boldsymbol{\beta}}_{\text{WGEE}}, \hat{\boldsymbol{\gamma}}_{jd}^1)} + \frac{1}{T} \text{E} \{ \sum_{j=1}^T \boldsymbol{S}_{ij}^1(\boldsymbol{\beta}_0, \boldsymbol{\gamma}_{jd,0}^1) \} + o_p(1)$$

$$\xrightarrow{p} \frac{1}{\sum_{j=1}^T P(R_{ij}=1)} \text{E} \left[\sum_{j=1}^T \frac{R_{ij} \{ \boldsymbol{S}_{ij}(\boldsymbol{\beta}_0) - \boldsymbol{S}_{ij}^1(\boldsymbol{\beta}_0, \boldsymbol{\gamma}_{jd,0}^1) \}}{1 + \hat{\boldsymbol{\lambda}}^T \boldsymbol{g}_j(\boldsymbol{\alpha}_*, \boldsymbol{\beta}_{\text{GEE},*}, \boldsymbol{\gamma}_{jd,*})} \right]$$

$$= \frac{1}{\sum_{j=1}^T P(R_{ij}=1)} \text{E} \left[\text{E} \left\{ \sum_{j=1}^T \frac{R_{ij} \{ \boldsymbol{S}_{ij}(\boldsymbol{\beta}_0) - \boldsymbol{S}_{ij}^1(\boldsymbol{\beta}_0, \boldsymbol{\gamma}_{jd,0}^1) \}}{1 + \hat{\boldsymbol{\lambda}}^T \boldsymbol{g}_j(\boldsymbol{\alpha}_*, \boldsymbol{\beta}_{\text{GEE},*}, \boldsymbol{\gamma}_{jd,*})} \middle| R_{ij}, \mathbf{X}_{ij}, \bar{\mathbf{Z}}_{i,(j-1)} \right\} \right] = \mathbf{0}.$$

Appendix C: Similarity Between EL wGEE and AIPW

When one missingness model and one imputation model are specified, and both models are correctly specified, the proposed estimator is similar to the AIPW estimator.

From equation (8),

$$\mathbf{S}(\boldsymbol{\beta}_{\text{EL}}) = \sum_{i=1}^N \frac{\partial \{\boldsymbol{\mu}_i(\boldsymbol{\beta}_0)\}^T}{\partial \boldsymbol{\beta}_0} (\mathbf{V}_i)^{-1} \mathbf{W}_{\text{EL},i} \{\mathbf{Y}_i - \boldsymbol{\mu}_i(\boldsymbol{\beta}_0)\} = \mathbf{0},$$

$$\sum_{i=1}^N \sum_{j=1}^{n_i} w_{ij} \mathbf{S}_{ij}(\boldsymbol{\beta}_0) = \mathbf{0}.$$

By using equation (5), $\sum_{i=1}^N \sum_{j=1}^{n_i} w_{ij} \mathbf{S}_{ij}^1(\hat{\boldsymbol{\beta}}_{\text{GEE}}^1, \hat{\boldsymbol{\gamma}}_{jd}^1) = \sum_{i=1}^N \sum_{j=1}^T \mathbf{S}_{ij}^1(\hat{\boldsymbol{\beta}}_{\text{GEE}}^1, \hat{\boldsymbol{\gamma}}_{jd}^1)/(NT)$,

$$\sum_{i=1}^N \sum_{j=1}^{n_i} \hat{w}_{ij} \mathbf{S}_{ij}(\boldsymbol{\beta}_0) -$$

$$\sum_{i=1}^N \sum_{j=1}^{n_i} \hat{w}_{ij} \mathbf{S}_{ij}^1(\hat{\boldsymbol{\beta}}_{\text{GEE}}^1, \hat{\boldsymbol{\gamma}}_{jd}^1) + \sum_{i=1}^N \sum_{j=1}^T \mathbf{S}_{ij}^1(\hat{\boldsymbol{\beta}}_{\text{GEE}}^1, \hat{\boldsymbol{\gamma}}_{jd}^1)/(NT) = \mathbf{0}.$$

When the missingness model is correctly specified, from equation (A.2),

$$\hat{w}_{ij} = \frac{1}{m} \frac{\theta^1(\hat{\boldsymbol{\alpha}}^1, \hat{\boldsymbol{\beta}}_{\text{wGEE}}^1)/\pi_{ij}^1(\hat{\boldsymbol{\alpha}}^1)}{1 + \hat{\boldsymbol{\kappa}}^T \hat{\boldsymbol{g}}_{ij}(\hat{\boldsymbol{\alpha}}, \hat{\boldsymbol{\beta}}_{\text{GEE}}, \hat{\boldsymbol{\beta}}_{\text{wGEE}}, \hat{\boldsymbol{\gamma}}_{jd})/\pi_{ij}^1(\hat{\boldsymbol{\alpha}}^1)} \xrightarrow{p} \frac{1}{m} \frac{\theta^1(\hat{\boldsymbol{\alpha}}^1, \hat{\boldsymbol{\beta}}_{\text{wGEE}}^1)}{\pi_{ij}^1(\hat{\boldsymbol{\alpha}}^1)},$$

thus,

$$\frac{1}{m} \sum_{i=1}^N \sum_{j=1}^T \frac{\theta^1(\hat{\boldsymbol{\alpha}}^1, \hat{\boldsymbol{\beta}}_{\text{wGEE}}^1) R_{ij}}{\pi_{ij}^1(\hat{\boldsymbol{\alpha}}^1)} \mathbf{S}_{ij}(\boldsymbol{\beta}_0)$$

$$- \frac{1}{m} \sum_{i=1}^N \sum_{j=1}^T \frac{\theta^1(\hat{\boldsymbol{\alpha}}^1, \hat{\boldsymbol{\beta}}_{\text{wGEE}}^1) R_{ij}}{\pi_{ij}^1(\hat{\boldsymbol{\alpha}}^1)} \mathbf{S}_{ij}^1(\hat{\boldsymbol{\beta}}_{\text{GEE}}^1, \hat{\boldsymbol{\gamma}}_{jd}^1) + \sum_{i=1}^N \sum_{j=1}^T \mathbf{S}_{ij}^1(\hat{\boldsymbol{\beta}}_{\text{GEE}}^1, \hat{\boldsymbol{\gamma}}_{jd}^1)/(NT) = \mathbf{0}.$$

By using $y_{ij} \perp R_{ij} | (\mathbf{X}_{ij}, \bar{\mathbf{Z}}_{i,(j-1)})$ and $\theta^1(\hat{\boldsymbol{\alpha}}^1, \hat{\boldsymbol{\beta}}_{\text{wGEE}}^1) = \sum_{i=1}^N \sum_{j=1}^T \pi_{ij}^s(\hat{\boldsymbol{\alpha}}^s, \hat{\boldsymbol{\beta}}_{\text{wGEE}}^s)/(NT)$,

$$\frac{\theta^1(\hat{\boldsymbol{\alpha}}^1, \hat{\boldsymbol{\beta}}_{\text{wGEE}}^1)}{m} \sum_{i=1}^N \sum_{j=1}^T \frac{R_{ij}}{\pi_{ij}^1(\hat{\boldsymbol{\alpha}}^1)} \mathbf{S}_{ij}(\boldsymbol{\beta}_0)$$

$$- \frac{\theta^1(\hat{\boldsymbol{\alpha}}^1, \hat{\boldsymbol{\beta}}_{\text{wGEE}}^1)}{m} \sum_{i=1}^N \sum_{j=1}^T \frac{R_{ij} - \frac{m}{\theta^1(\hat{\boldsymbol{\alpha}}^1, \hat{\boldsymbol{\beta}}_{\text{wGEE}}^1) NT} \pi_{ij}^1(\hat{\boldsymbol{\alpha}}^1)}{\pi_{ij}^1(\hat{\boldsymbol{\alpha}}^1)} \mathbf{S}_{ij}^1(\hat{\boldsymbol{\beta}}_{\text{GEE}}^1, \hat{\boldsymbol{\gamma}}_{jd}^1) = \mathbf{0},$$

$$\frac{\theta^1(\hat{\boldsymbol{\alpha}}^1, \hat{\boldsymbol{\beta}}_{\text{wGEE}}^1)}{m} \sum_{i=1}^N \sum_{j=1}^T \frac{R_{ij}}{\pi_{ij}^1(\hat{\boldsymbol{\alpha}}^1)} \mathbf{S}_{ij}(\boldsymbol{\beta}_0)$$

$$- \frac{\theta^1(\hat{\alpha}^1, \hat{\beta}_{\text{WGEE}}^1)}{m} \sum_{i=1}^N \sum_{j=1}^T \frac{R_{ij} - \frac{m}{\sum_{i=1}^N \sum_{j=1}^T \pi_{ij}^s(\hat{\alpha}^s, \hat{\beta}_{\text{WGEE}}^s)} \pi_{ij}^1(\hat{\alpha}^1)}{\pi_{ij}^1(\hat{\alpha}^1)} \mathbf{S}_{ij}^1(\hat{\beta}_{\text{GEE}}^1, \hat{\mathbf{Y}}_{jd}^1) + o_p(1).$$

When the imputation model is correctly specified, $\hat{\beta}_{\text{GEE}}^1 \xrightarrow{p} \beta_0$, thus,

$$\begin{aligned} &\xrightarrow{p} \frac{\theta^1(\hat{\alpha}^1, \hat{\beta}_{\text{WGEE}}^1)}{m} \sum_{i=1}^N \sum_{j=1}^T \frac{R_{ij}}{\pi_{ij}^1(\hat{\alpha}^1)} \mathbf{S}_{ij}(\beta_0) \\ &\quad - \frac{\theta^1(\hat{\alpha}^1, \hat{\beta}_{\text{WGEE}}^1)}{m} \sum_{i=1}^N \sum_{j=1}^T \frac{R_{ij} - \frac{m}{\sum_{i=1}^N \sum_{j=1}^T \pi_{ij}^s(\hat{\alpha}^s, \hat{\beta}_{\text{WGEE}}^s)} \pi_{ij}^1(\hat{\alpha}^1)}{\pi_{ij}^1(\hat{\alpha}^1)} \mathbf{S}_{ij}^1(\beta_0, \hat{\mathbf{Y}}_{jd}^1), \end{aligned}$$

in the probability as $N \rightarrow \infty$, since $\sum_{i=1}^N \sum_{j=1}^T \pi_{ij}^s(\hat{\alpha}^s, \hat{\beta}_{\text{WGEE}}^s) \rightarrow m$; thus,

$$\frac{\theta^1(\hat{\alpha}^1, \hat{\beta}_{\text{WGEE}}^1)}{m} \sum_{i=1}^N \sum_{j=1}^T \frac{R_{ij}}{\pi_{ij}^1(\hat{\alpha}^1)} \mathbf{S}_{ij}(\beta_0) - \frac{\theta^1(\hat{\alpha}^1, \hat{\beta}_{\text{WGEE}}^1)}{m} \sum_{i=1}^N \sum_{j=1}^T \frac{R_{ij} - \pi_{ij}^1(\hat{\alpha}^1)}{\pi_{ij}^1(\hat{\alpha}^1)} \mathbf{S}_{ij}^1(\beta_0, \hat{\mathbf{Y}}_{jd}^1) + o_p(1)$$

$$\xrightarrow{p} \frac{\theta^1(\hat{\alpha}^1, \hat{\beta}_{\text{WGEE}}^1)}{\sum_{j=1}^T P(R_{ij}=1)} E \left\{ \sum_{j=1}^T \frac{R_{ij}}{\pi_{ij}^1(\hat{\alpha}^1)} \mathbf{S}_{ij}(\beta_0) - \sum_{j=1}^T \frac{R_{ij} - \pi_{ij}^1(\hat{\alpha}^1)}{\pi_{ij}^1(\hat{\alpha}^1)} \mathbf{S}_{ij}^1(\beta_0, \hat{\mathbf{Y}}_{jd}^1) \right\} = \mathbf{0}. \quad (\text{C.1})$$

Equation (C.1) is similar to equation (4) of the AIPW described in Han et al. (2015).

Tables

Table 1. Parameter estimates $\hat{\beta}_{\text{TRT}}$ in the plaque psoriasis study.

	$\hat{\pi}^1$	$\hat{\pi}^2$	$\hat{\phi}^1$	$\hat{\phi}^2$	$\hat{\beta}_{\text{TRT}}$	SE	Odds ratio	95% confidence interval
CLUDE	-	-	-	-	1.531	0.283	4.62	2.653–8.055
GEE	-	-	-	-	1.452	0.280	4.27	2.468–7.398
wGEE	1	0	0	0	1.402	0.278	4.06	2.357–7.004
	0	1	0	0	1.417	0.282	4.12	2.372–7.168
AIPW	1	0	1	0	1.399	0.274	4.05	2.369–6.930
	0	1	0	1	1.367	0.280	3.93	2.269–6.790
EL wGEE	1	1	1	1	1.367	0.281	3.92	2.262–6.799
	1	0	1	0	1.407	0.274	4.08	2.384–6.988
	0	1	0	1	1.343	0.280	3.83	2.212–6.630

CLUDE: logistic regression; wGEE: weighted generalized estimating equations; AIPW: augmented inverse probability weighted complete-case estimating equations (Seaman and Copas 2009); EL wGEE: proposed estimator; SE: standard error.

The names of the estimators have the form “method-0000,” with each digit of the four-digit number, from left to right, indicating whether or not $\pi_{ij}^1(\alpha^1)$, $\pi_{ij}^2(\alpha^2)$, $\phi_{ija}^1(\gamma_{ja}^1)$, or $\phi_{ija}^2(\gamma_{ja}^2)$ is included, where “1” indicates “included” and “0” indicates “not included.” In the results of the proposed method and AIPW method, only two or three models are shown; EL wGEE-1111 for the multiply robust model and EL wGEE-1010, and EL wGEE-0101 (or AIPW-1010, AIPW-0101) for the doubly robust model

Figures

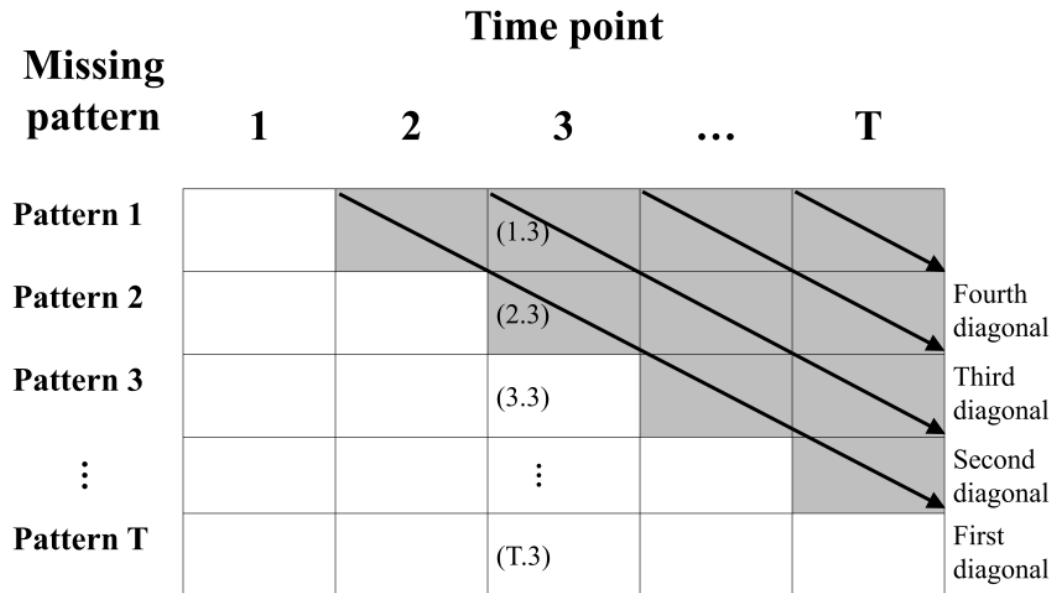


Figure 1. Imputation method of Paik.

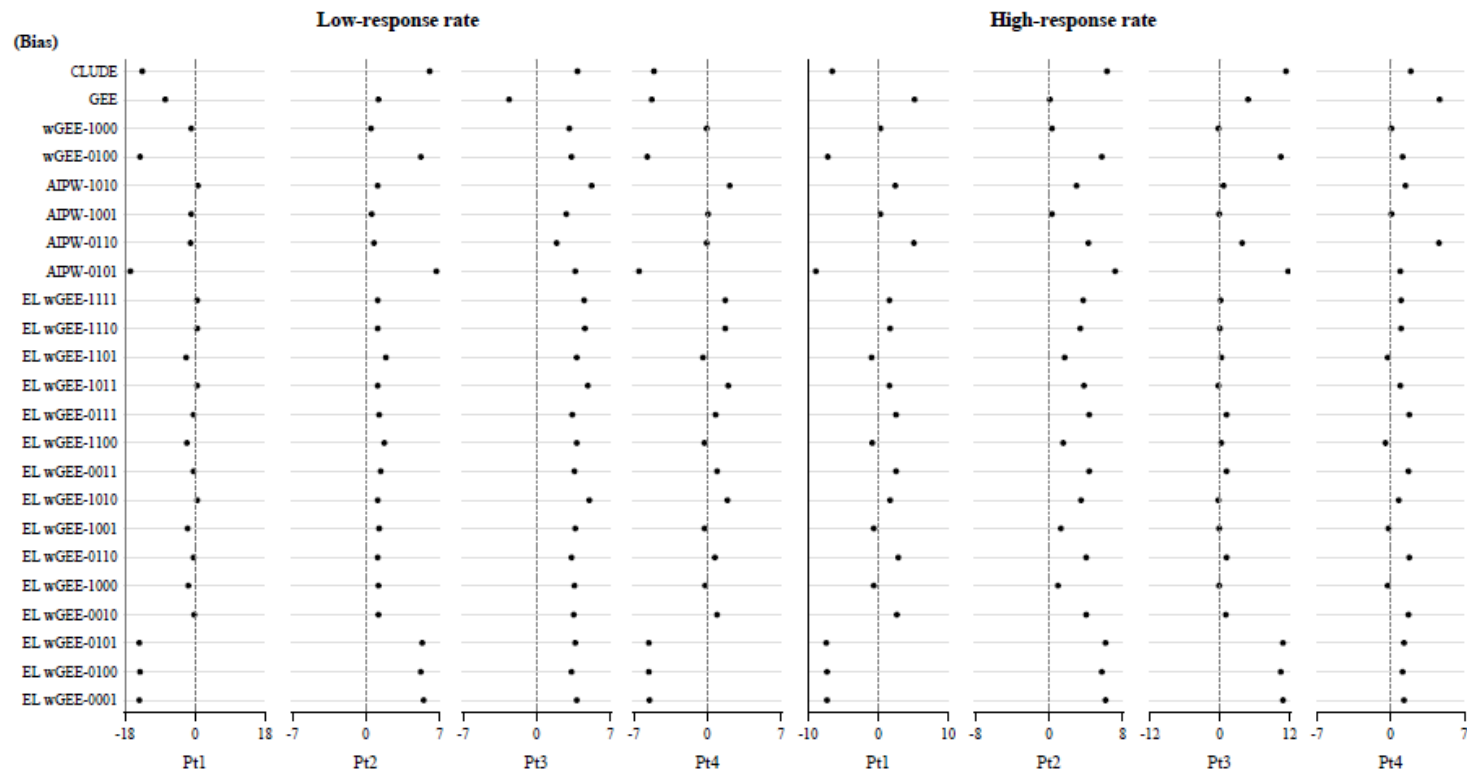


Figure 2. Bias of $\hat{\beta}_{\text{TRT}}$ in each estimator obtained using 1000 Monte Carlo datasets ($n = 150$ in each treatment group; proportion of missing data at $j=4$ is 30%). CLUDE: logistic regression; wGEE: weighted generalized estimating equations; AIPW: augmented inverse probability weighted complete-case estimating equations (Seaman and Copas 2009); EL wGEE: proposed estimator. The results were multiplied by 100 to reduce the number of

decimal places. The notation of these estimators in the table includes the names of the estimators and the model specification, which have the form “method-0000.” Each digit in the four-digit number, from left to right, indicates whether or not $\pi_{ij}^1(\boldsymbol{\alpha}^1)$, $\pi_{ij}^2(\boldsymbol{\alpha}^2)$, $\phi_{ija}^1(\boldsymbol{\gamma}_{ja}^1)$, or $\phi_{ija}^2(\boldsymbol{\gamma}_{ja}^2)$ is included, where “1” indicates “included” and “0” indicates “not included.” When estimating the parameters of the missingness/imputation models, quasi-complete separation occurred in some of the simulation datasets. Those datasets were excluded from the analysis. The detailed values are shown in Web Appendix B of Supporting information.

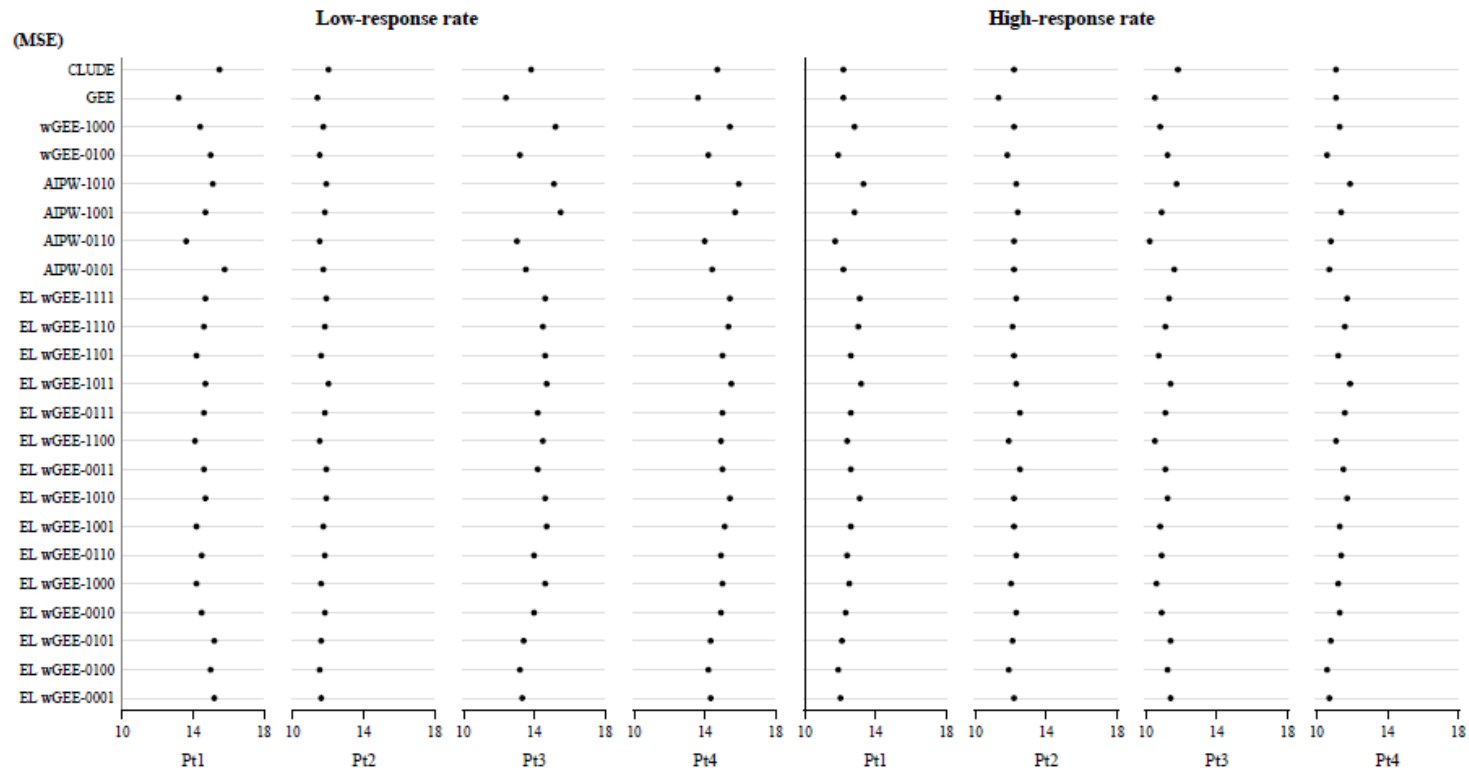


Figure 3. Mean square errors (MSE) of $\hat{\beta}_{\text{TRT}}$ in each estimator obtained using 1000 Monte Carlo datasets ($n = 150$ in each treatment group; proportion of missing data at $j=4$ is 30%). CLUDE: logistic regression; wGEE: weighted generalized estimating equations; AIPW: augmented inverse probability weighted complete-case estimating

equations (Seaman and Copas 2009); EL wGEE: proposed estimator. The results were multiplied by 100 to reduce the number of decimal places. The detailed values are shown in Web Appendix B of Supporting information.

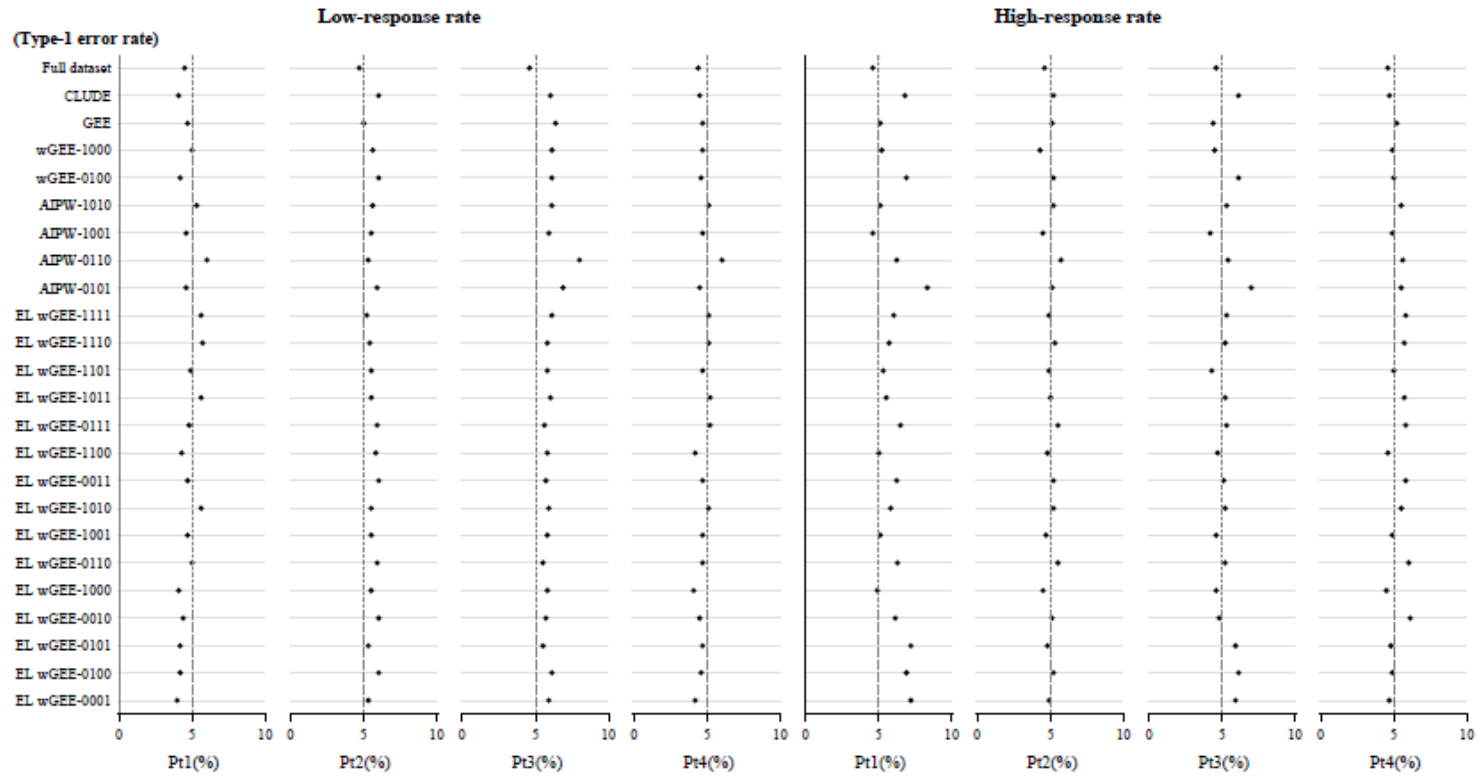


Figure 4. Type-1 error rates in each estimator obtained using 1000 Monte Carlo datasets ($n = 150$ in each treatment group; proportion of missing data at $j=4$ is 30%). Full dataset: logistic regression with full dataset; CLUDE: logistic regression; wGEE: weighted generalized estimating equations; AIPW: augmented inverse probability weighted

complete-case estimating equations (Seaman and Copas 2009); EL wGEE: proposed estimator. The results were multiplied by 100 to reduce the number of decimal places. The detailed values are shown in Web Appendix B of Supporting information