

AN EFFICIENT DOUBLY-ROBUST IMPUTATION FRAMEWORK FOR LONGITUDINAL DROPOUT, WITH AN APPLICATION TO AN ALZHEIMER'S CLINICAL TRIAL

BY YUQI QIU^{1,a} AND KAREN MESSER^{2,b}

¹*School of Statistics, East China Normal University, yqqiu@fem.ecnu.edu.cn*

²*Division of Biostatistics and Bioinformatics, University of California San Diego, kmesser@health.ucsd.edu*

We develop a novel doubly-robust (DR) imputation framework for longitudinal studies with monotone dropout, motivated by the informative dropout that is common in FDA-regulated trials for Alzheimer's disease. In this approach the missing data are first imputed using a doubly-robust augmented inverse probability weighting (AIPW) estimator; then the imputed completed data are substituted into a full-data estimating equation, and the estimate is obtained using standard software. The imputed completed data may be inspected and compared to the observed data, and standard model diagnostics are available. The same imputed completed data can be used for several different estimands, such as subgroup analyses in a clinical trial, allowing for reduced computation and increased consistency across analyses. We present two specific DR imputation estimators, AIPW-I and AIPW-S, study their theoretical properties, and investigate their performance by simulation. AIPW-S has substantially reduced computational burden, compared to many other DR estimators, at the cost of some loss of efficiency and the requirement of stronger assumptions. Simulation studies support the theoretical properties and good performance of the DR imputation framework. Importantly, we demonstrate their ability to address time-varying covariates, such as a time by treatment interaction. We illustrate using data from a large randomized Phase III trial, investigating the effect of donepezil in Alzheimer's disease, from the Alzheimer's Disease Cooperative Study (ADCS) group.

1. Background and a motivating example. Dropout rates of 25% or more are common in large randomized trials for Alzheimer's disease and other dementias. The dropout rate is generally higher for patients with more severe disease and also for patients receiving active treatment (Figure 1). Thus, estimates of treatment effect from the trial may be substantially biased, unless the dropout is properly accounted for. However, the primary analysis of such a trial is tightly prespecified. U.S. Food and Drug Administration (FDA) guidance and practices support the use of a mixed-effects model with repeated measures (MMRM) or a generalized estimating equations (GEE) approach with a restricted set of covariates; the primary analysis usually tests a model-adjusted estimate of a contrast between treatment arms. Because of restricted covariates and modeling strategies, it is very possible that the covariate adjustment may be insufficient to assure unbiased estimates from the model. To address this problem in the regulatory setting of clinical trials, additional imputation-based sensitivity analyses are recommended to assess the potential bias from dropout, usually through sequential multiple imputation by chained equations (International Council for Harmonisation Of Technical Requirements for Pharmaceuticals for Human Use (2017), Committee for Medicinal Products for Human Use (2010)).

Motivated by this setting, we investigate a principled doubly-robust approach to such imputation for longitudinal dropout, derived from the theory of optimal augmented inverse

Received November 2021; revised November 2022.

Key words and phrases. Doubly-robust estimator, monotone dropouts, imputation methods, Alzheimer's disease, randomized trials.

probability weighted (AIPW) estimators, as developed by Bang and Robins (2005), Tsiatis (2006), Seaman and Copas (2009), Tsiatis, Davidian and Cao (2011), Rotnitzky et al. (2012), Schnitzer, Lok and Bosch (2016), and others in the longitudinal setting. Formally, the interest lies in obtaining a robust, consistent, and asymptotically normally distributed (CAN) estimate of an estimand β , defined as the solution to a prespecified estimating equation $E[U(\beta)] = 0$. A simple example is where $\beta = E(Y)$ and $U(\beta) = Y - \beta$. If $U(\beta)$ is applied to the “full” data (fully observed data with no dropout), the solution $\hat{\beta}$ is a CAN estimator of β . However, the primary intent-to-treat analysis from the trial usually applies $U(\beta)$ to the observed data. Under a missing at random (MAR) assumption conditional on covariates, this would indeed provide a consistent estimator if all covariates are appropriately controlled for. As explained above, however, in our setting U is usually a GEE derived from a generalized linear model with limited covariates rather than a fully specified likelihood. Therefore, the approach, in practice, is generally not assumed to provide unbiased estimates in the presence of missing data.

Consistent estimators in the setting of longitudinal monotone dropout have been well studied in theory, including doubly-robust (DR) estimators; for a recent review, see Seaman and Vansteelandt (2018). Paik (1997) constructed an imputation approach to obtain consistent estimators for longitudinal data with dropout, using a sequential regression algorithm to impute the missing outcome values. Robins, Rotnitzky and Zhao (1995) noted that consistent estimates of β may be obtained by solving an inverse probability weighted (IPW) version of the estimating functions $U(\beta)$ applied to the data from completers only, or by using an AIPW estimator, which increases the efficiency by augmenting the IPW estimator with the observed information from dropout subjects using a regression model. Scharfstein, Rotnitzky and Robins (1999) noticed that these AIPW estimators are doubly-robust (DR), in that they are consistent when either the regression model or the IPW model is correctly specified. Bang and Robins (2005) proposed a regression representation of an AIPW estimator, using a recursive regression approach which incorporates the IPW’s as covariates; however, in the longitudinal setting, they only explicitly studied the case where $\beta = E[Y]$ at last visit. Tsiatis (2006) developed the theory of optimal AIPW estimating equations in the setting of longitudinal data with monotone dropout for a general class of $U(\beta)$ and implemented an improved DR estimator with MMRM as the regression model in Tsiatis, Davidian and Cao (2011), which, however, is quite complex in practice. Following Tsiatis (2006), Seaman and Copas (2009) considered $U(\beta)$, arising as generalized estimating equations, and used Paik’s sequential regression framework to construct an AIPW estimator for the regression coefficients. Rotnitzky et al. (2012) further developed an optimal AIPW estimator which may obtain good efficiency when the outcome model is misspecified and provided simulation studies in the cross-sectional setting. More recently, Schnitzer, Lok and Bosch (2016) adapted the Bang and Robins (2005) approach in the longitudinal setting to include more general estimands defined by a GLM with baseline covariates; however, the class of estimands, in particular, does not include the regression coefficient of a time-varying covariate. Schnitzer, Lok and Bosch (2016) also study a closely related targeted maximum likelihood estimator (TMLE) (van der Laan and Rubin (2006)) and show that it has similar performance as the adapted Bang and Robins method by simulation. Long, Hsu and Li (2012) and Hsu et al. (2016) considered incorporating a DR estimator into a multiple imputation (MI) approach in the cross-sectional setting; however, this differs from the longitudinal approach we consider here.

It is worth noting that, compared with the cross-sectional setting, DR methods for longitudinal data are comparatively less well developed and less often used in practice. In particular, we know of relatively few simulation studies that study these longitudinal doubly-robust estimators (Seaman and Copas (2009), Tsiatis, Davidian and Cao (2011), Schnitzer, Lok and Bosch (2016)). While several approaches can in theory estimate the coefficient of a time varying covariate (Seaman and Copas (2009), Tsiatis, Davidian and Cao (2011)), we did not find

simulation studies that cover this case which is critical to the clinical trials setting. We are unfamiliar with examples of the use of DR estimators in clinical trials, where we think they may have a useful and important role.

In this paper we further develop results from [Seaman and Copas \(2009\)](#) into a general imputation-based framework for construction of a locally efficient doubly-robust estimator in the setting of longitudinal data with monotone dropout. The approach is to impute a complete dataset, using a suitable doubly-robust estimator, and then apply the estimating functions $U(\beta)$ to the fully-imputed data. We first investigate a standard AIPW-based doubly-robust estimator (AIPW-I) within the framework, and then we propose a simpler doubly-robust estimator (AIPW-S) also within the framework. Confidence intervals are provided by the bootstrap in an imputation setting which is familiar to practitioners of clinical trials. This imputation framework has several advantages compared to existing approaches in the literature: (1) once the imputed completed data are obtained, the same completed dataset will support doubly-robust estimation of additional estimands, such as additional additional group or subgroup effects, providing computational efficiency and consistency across multiple analyses; (2) the imputed completed data can be inspected, using standard descriptive statistics, to better understand the behavior of the resulting doubly-robust estimator and any differences from the primary analysis (which uses the observed data directly). Both the imputation step and the analysis step use well-understood modeling approaches which are easy to prespecify and are suitable for the clinical trials setting. The imputation framework is applicable to any AIPW-based doubly-robust estimator, including the existing highly-developed doubly-robust methods, and we hope will facilitate the construction of doubly robust estimates of causal effects in clinical trial settings. Importantly, our proposed framework supports doubly robust estimation of the coefficient of a time-varying covariate as the estimand, such as a treatment by time interaction term.

The outline of the approach is as follows: for $U(\beta)$ a GEE, [Seaman and Copas \(2009\)](#) constructed a DR estimator of regression coefficients by substituting Paik's sequential regression into the optimal longitudinal AIPW equations of [Tsiatis \(2006\)](#) and solving the resulting equations by Newton–Raphson. Our contribution is to notice that the resulting estimating equations might be rearranged into an imputation form, after establishing some algebraic identities, and to exploit the result. The resulting estimator only requires standard software tools and inherits the optimality properties of the doubly-robust imputation for a wide class of estimating functions. Using this framework, we develop a computationally simpler doubly-robust estimator, AIPW-S, under explicit assumptions which may be practical in clinical trials and other settings. Finally, we study the performance of these estimators through simulation, and through an application to the primary estimands of interest in a clinical trial of prodromal Alzheimer's Disease. These simulation studies may be of independent interest, as they add to the sparse literature on longitudinal AIPW approaches.

1.1. Organization of this paper. In Section 2 we describe our motivating example. Section 3 gives notation and the details of consistent IPW and sequential regression approaches for longitudinal data with dropout. These form the building blocks of our DR imputation method. We briefly review existing DR methods for longitudinal data in Section 4. In Section 5 we develop the AIPW-based DR imputation framework for general longitudinal data estimating equations. In Section 6 we use the DR imputation framework to define two specific DR estimators for longitudinal data with dropout, AIPW-I and AIPW-S. We use simulation to compare AIPW-I and AIPW-S with the original Bang and Robins estimator as well as with maximum likelihood and GEE approaches in Section 7. Section 8 presents an application to the donepezil trial in Alzheimer's disease, and Section 9 is discussion and conclusions.

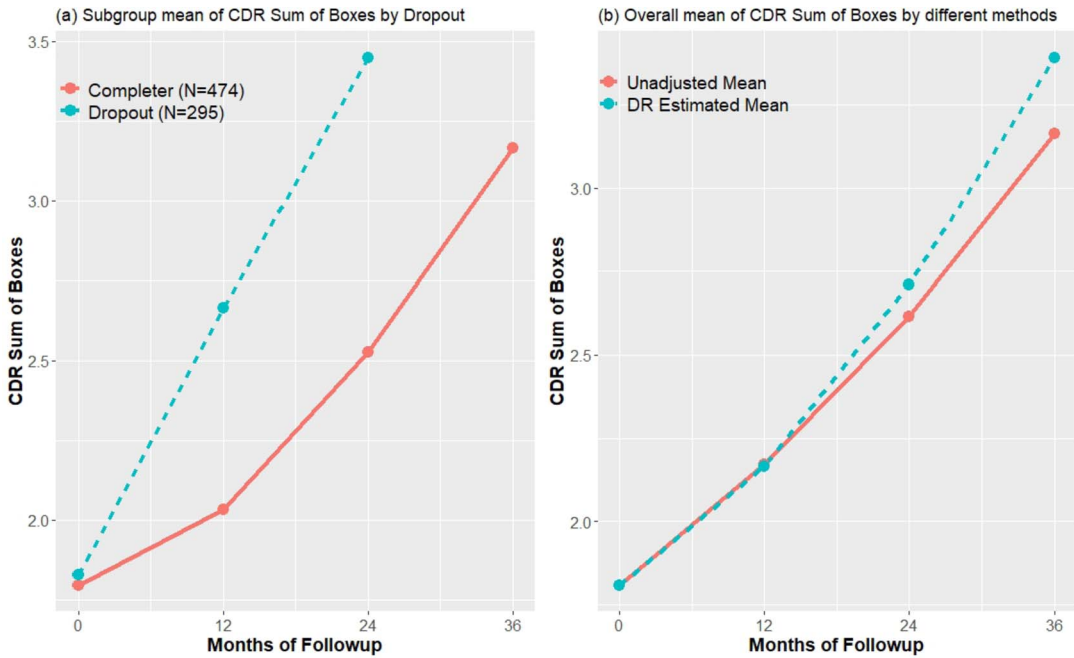


FIG. 1. Time course of the mean Clinical Dementia Rating Sum of Boxes (CDR-SOB) score, using all subjects who provided data on this measure in the MCI trial ($n = 769$). The CDR-SOB is commonly used in Alzheimer's disease to stage severity of dementia, with a higher score indicating worse severity; it is expected to increase over time in this population with mild cognitive impairment (MCI) at baseline. Left panel: Mean CDR-SOB at each time point, for subjects who completed the study (solid line, 62% of subjects) vs. subjects who dropped out before the end of the study (dashed line, 38% of subjects). Subjects who dropped out are observed to have higher change scores. Right panel: Mean CDR-SOB using the available data at each time point (solid line); doubly-robust estimate of the mean CDR-SOB, using all available data and adjusting for dropout (dashed line).

2. A motivating example: The MCI trial of donepezil. Donepezil is a widely used cholinesterase inhibitor that improves symptoms and may delay the clinical diagnosis of Alzheimer's disease (AD) in subjects with the amnesic form of mild cognitive impairment (MCI). A randomized, double-blind, placebo-controlled, parallel-arm trial was conducted by the Alzheimer's Disease Cooperative Study (ADCS) between March 1999 and January 2004 (Petersen et al. (2005)). The study compared the time to progression from MCI to possible or probable AD among 769 subjects with MCI who were randomized to receive donepezil ($n = 253$), vitamin E ($n = 257$), or placebo ($n = 259$) for 36 months. Final dropout rates were 42.7%, 38.1%, and 32.0% for the donepezil, Vitamin E, and placebo arms, respectively. Figure 1 displays evidence of bias due to dropout for the mean score on the Clinical Dementia Rating sum of boxes (CDR-SOB), one of the common primary outcome measures of AD trials. The left panel shows that patients who eventually withdrew from the study had a much higher CDR-SOB (higher is worse) than the completers, indicating that those who dropped out had greater cognitive impairment and that this gap increased over time. The figure on the right shows that the estimated mean from a DR approach is consistently higher than the mean of the observed data, especially at the later times with more dropout, indicating that a DR approach may help to improve estimated effects from this trial. These data are taken from the data archives at the ADCS; these and similar data are often used in simulation studies that inform the design of current AD trials.

3. Regression modeling and inverse probability weighting approaches to longitudinal dropout.

3.1. *Notation and data structure.* Assume we have N i.i.d. subjects potentially observed at times $j = 1, \dots, M$, and for individual i at time j , there is data $L_{ij} = (Y_{ij}, X_{ij})$, where Y_{ij} is a univariate outcome and X_{ij} is a vector of potentially time-varying covariates; when X_{ij} is independent of time, it will be simplified to a vector of always observed baseline covariates X_{i0} for any j . Let $\bar{L}_{ij} = (L_{i1}^T, \dots, L_{ij}^T)^T$ denote the historical data from time 1 to j . We may drop the subscript i when the meaning is clear. We assume the distribution of \bar{L}_{iM} has finite second moments.

Each subject can potentially drop out from the study. Let $R_{ij} \in \{0, 1\}$ be a missing indicator so that we observe R_{ij} and $(R_{ij}Y_{ij}, R_{ij}X_{ij})$ at time j . Under the assumption of monotone dropout, if $R_j = 0$, then, for any $t > j$, $R_t = 0$. Let C_{ij} be a censoring indicator, where $C_{ij} = 1$ indicates j is the final observed time for subject i , otherwise $C_{ij} = 0$; let J_i be the index of the last observed time point for subject i so that $C_{i,J_i} = 1$ for subject i . For completers, define $J_i = M$ and $C_M = R_M = 1$. Under the missing completely at random assumption (MCAR), R_j is independent of \bar{L}_M . Under the missing at random assumption (MAR), $\Pr(R_j = 0 | \bar{L}_M, R_{j-1} = 1) = \Pr(R_j = 0 | \bar{L}_{j-1}, R_{j-1} = 1)$ so that the probability of a missing outcome depends only on previously observed data. We also assume there is probability bounded away from zero of seeing full data over the whole support of \bar{L}_{iM} : $\Pr(R_M = 1 | \bar{L}_M) > \epsilon > 0$ for some existing ϵ .

3.2. *Estimating equations which define the estimand.* We assume that there is a vector of parameters β and a corresponding vector of sufficiently smooth estimating functions $U(\cdot, \beta)$ such that β^* is the unique solution (the truth) to $E[U(\bar{L}_{iM}, \beta)] = 0$. One of the parameters in β^* is the primary estimand of interest in the study. The solution $\hat{\beta}$ to the full data, estimating equations $\sum_{i=1}^N U_i(\bar{L}_{iM}, \beta) = 0$, is consistent for β^* and asymptotically normal, by standard arguments. It is often assumed that the data follow a generalized linear model for the mean $\mu_j = E(Y_j | X_j)$, with link function $g(\mu_j) = X_j \beta$, including of course the MMRM of the Introduction. Thus, $U(\bar{L}_{iM}, \beta)$ might be taken to be the score equations from the likelihood or, alternatively, a set of generalized estimating equations (GEE) (Liang and Zeger (1986)) applied to the full data,

$$(1) \quad \sum_{i=1}^N U_i(\bar{L}_{iM}) = \sum_{i=1}^N \frac{\partial \mu_i^T}{\partial \beta} V_i^{-1} (\mathbf{Y}_i - \mu_i) = 0,$$

where $\mathbf{Y}_i = (Y_{i1}, \dots, Y_{iM})^T$, $\mu_i = (\mu_{i1}, \dots, \mu_{iM})^T$, and V_i is an assumed working covariance matrix for \mathbf{Y}_i . Here the efficient choice for V_i^{-1} is the true covariance matrix of the data. However, under general regularity conditions the solution $\hat{\beta}$ to the full data GEE's (equation (1)) is consistent for β^* and asymptotically normal for any arbitrary V_i .

With missing data, instead of equation (1) we observe

$$(2) \quad \sum_{i=1}^N U_i(\bar{L}_{iJ}) = \sum_{i=1}^N \frac{\partial \mu_i^T}{\partial \beta} V_i^{-1} D_i (\mathbf{Y}_i - \mu_i) = 0,$$

where $D_i = \text{diag}(R_{i1}, \dots, R_{iM})$. If the data are MCAR, the solution $\hat{\beta}$ to equation (2) remains consistent, because $E[R_j Y_j] = E[Y_j]$, and thus equation (2) is a consistent estimator of $E[U_i(\bar{L}_{iM})]$. However, when the dropout is MAR such that $E[R_j Y_j] \neq E[Y_j]$, the solution $\hat{\beta}$ to equation (2) will be consistent for β if the U are the score equations from the correct likelihood but not generally otherwise.

3.3. *IPW estimating equations for dropout that is MAR.* Robins, Rotnitzky and Zhao (1995) showed how to incorporate inverse-probability weights into U applied to observed data when the dropout is MAR. Let

$$(3) \quad \lambda_{ij} = P(R_{ij} = 0 | R_{i,j-1} = 1, \bar{L}_{i,j-1})$$

be the discrete-time hazard function of i at time j , let

$$(4) \quad \pi_{ij} = \prod_{t=1}^j (1 - \lambda_{it}) = P(R_{ij} = 1 | \bar{L}_{i,j-1})$$

be the corresponding probability of being observed at time j , and let the weight matrix be $W_i = \text{diag}(R_{i1}/\pi_{i1}, \dots, R_{iM}/\pi_{iM})$. The hazard function can be consistently estimated by logistic regression if the MAR assumption holds; thus, π_j and W_π can be consistently estimated as well. Then the inverse probability weighted GEE (WGEE) has the estimating equations

$$(5) \quad \sum_{i=1}^N U_i^W(\bar{L}_{iJ}) = \sum_{i=1}^N \frac{\partial \mu_i^T}{\partial \beta} V_i^{-1} W_i (\mathbf{Y}_i - \mu_i) = 0.$$

For longitudinal data with dropout that is MAR, WGEE provides consistent estimates of the parameters β^* if the π_{ij} 's are consistent, because then $E[R_j Y_j / \pi_j] = E[Y_j]$, even if the working correlation is not correctly specified.

3.4. *Regression-based sequential imputation, for dropout that is MAR.* Alternatively, Paik (1997) described a sequential regression approach for imputing the missing outcome values. Recall that J_i is the last observed visit for subject i so that Y_{ik} is missing for any $k > J_i$. In order to impute Y_{ik} , the idea is to conduct a recursive regression process by defining $k - 1$ parametric imputation models $m_k^j(\bar{L}_j) = E[Y_k | \bar{L}_j, R_{j+1} = 1]$, where j is taken from $k - 1$ to 1. By the MAR assumption, $E[Y_k | \bar{L}_j, R_{j+1} = 0] = E[Y_k | \bar{L}_j, R_{j+1} = 1]$. Hence, we may use observed data to construct a consistent estimated model $\hat{m}_k^j(\bar{L}_j)$ and then use the estimated model to impute missing Y_k recursively. This process can then be repeated sequentially for $k = 2, \dots, M$ in order to obtain a fully imputed data set.

Here we give a formal version of Paik's sequential mean imputation algorithm which will make the relation with Bang and Robins (2005) more explicit.

The above imputation requires $M(M - 1)/2$ estimated models. Then estimating equation (1) is solved, using the completed data,

$$(6) \quad \sum_{i=1}^N U_i(\hat{L}_{iM}^I) = \sum_{i=1}^N \frac{\partial \mu_i^T}{\partial \beta} V_i^{-1} (\hat{\mathbf{Y}}_i^I - \mu_i) = 0.$$

Under regularity conditions, as long as the mean models m_k^s are correctly specified and the data are MAR, this procedure gives consistent estimates of the parameters β^* (Paik (1997)).

4. Doubly-robust estimators for longitudinal data with dropout.

4.1. *Optimal longitudinal AIPW estimating equations.* In the longitudinal setting with known monotone MAR dropout mechanism, Tsiatis (2006) showed that, under regularity conditions, any consistent and asymptotically normal estimator $\hat{\beta}$ of β^* using the observed data solves AIPW estimating equations of the form

$$(7) \quad \sum_{i=1}^N \left(\frac{C_{i,M}}{\pi_{i,M}} U_i(\hat{\beta}, \bar{L}_{iM}) + \sum_{j=1}^{M-1} \left(\frac{C_{ij} - \lambda_{i,j+1} R_{ij}}{\pi_{i,j+1}} \right) H^j(\hat{\beta}, \bar{L}_{ij}) \right) = 0,$$

Algorithm 1: Paik's sequential mean imputation

Result: A fully imputed data \hat{Y}_{ij}^I , in which all longitudinal missing outcomes have been imputed.

1 For $k = 2, \dots, M$ sequentially:

2 Initialize: Identify all subjects i with $J_i \geq k$. Use these data to regress the observed values of Y_{ik} on $\bar{L}_{i,k-1}$ to obtain a consistently estimated model $\hat{m}_k^{k-1}(\bar{L}_{i,k-1})$. For subjects with $J_i \geq k - 1$, let

$$\hat{Y}_{ik}^{k-1} = \begin{cases} \hat{m}_k^{k-1}(\bar{L}_{i,k-1}) & \text{if } J = k - 1, \\ Y_{ik} & \text{if } J > k - 1. \end{cases}$$

3 For $s = k - 2, \dots, 1$ recursively: Identify all subjects i with $J_i \geq s + 1$, and regress the values of \hat{Y}_{ik}^{s+1} on \bar{L}_{is} to obtain a consistently estimated model $\hat{m}_k^s(\bar{L}_{is})$. For all subjects with $J_i \geq s$, let

$$\hat{Y}_{ik}^s = \begin{cases} \hat{m}_k^s(\bar{L}_{is}) & \text{if } J_i = s, \\ \hat{Y}_{ik}^{s+1} & \text{if } J_i > s. \end{cases}$$

4 Final step: Output the completed data $\hat{Y}_{ik}^I = \hat{Y}_{ik}^1$

where H^j is an arbitrary function. The choice

$$(8) \quad H^j(\boldsymbol{\beta}, \bar{L}_{ij}) = E(U(\boldsymbol{\beta}, \bar{L}_{iM}) | \bar{L}_{ij}, R_{ij} = 1)$$

yields the estimator with the smallest variance; thus (7) and (8) give the optimal observed data estimating equations. More generally, when λ_{ij} (and thus π_{ij}) is estimated by maximum likelihood and the \hat{H}^j in equation (8) are estimated by corresponding models, the solution $\hat{\boldsymbol{\beta}}^{\text{AIPW}}$ to the optimal estimating equations has the following properties (Tsiatis (2006)):

1. $\hat{\boldsymbol{\beta}}^{\text{AIPW}}$ is consistent for $\boldsymbol{\beta}^*$ and asymptotically normal if either the dropout models $\hat{\lambda}_j$ or the imputation models \hat{H}^j are correctly specified and thus consistent for the true conditional expectations they aim to estimate.

2. If both sets of models are correctly specified, $\hat{\boldsymbol{\beta}}^{\text{AIPW}}$ has the smallest asymptotic variance among all doubly-robust estimators of $\boldsymbol{\beta}$.

3. Improved, but more complex, doubly-robust estimators have been proposed, which also attain the minimum asymptotic variance when the imputation models are misspecified, but the dropout models are correctly specified and $\hat{\lambda}$ is an efficient estimator (Tsiatis, Davidian and Cao (2011)). We do not consider these estimators here.

4.2. *Seaman and Copas' doubly-robust AIPW estimator for longitudinal GEE.* For the case where $U_i(\boldsymbol{\beta})$ is a GEE of the form (1), Seaman and Copas (2009) proposed a two-step procedure: first, obtain estimates of the \hat{H}^j , and then substitute into the corresponding AIPW estimating equations (7). In particular, for subject i with $j \leq J_i$, take

$$(9) \quad \hat{H}^j(\hat{\boldsymbol{\beta}}, \bar{L}_{ij}) = \frac{\partial \mu_i^T}{\partial \hat{\boldsymbol{\beta}}} V_i^{-1} (\hat{\mathbf{Y}}_i^j - \mu_i),$$

where $\hat{\mathbf{Y}}_i^j$ is imputed from Paik's sequential mean imputation as in Algorithm 1. For $j > J_i$, H^j can be taken to be 0, as the weights for H^j in (7) are 0. Then the Newton–Raphson algorithm is used to solve equations (7) to obtain $\hat{\boldsymbol{\beta}}$.

Algorithm 2: Bang and Robins doubly-robust estimator $\hat{\mu}_{iM}^{BR}$ for $E(Y_{iM})$

Result: Bang and Robins doubly-robust estimator $\hat{\mu}_{iM}^{BR}$ for $E(Y_{iM})$ where

$$U_i(\bar{L}_{iM}) = Y_{iM} - E(Y_{iM})$$

- 1 **Preliminary step:** Estimate $\hat{\pi}_{i2}, \dots, \hat{\pi}_{iM}$ by maximum likelihood.
- 2 **Initialize:** Identify all subjects i with $R_{iM} = 1$. Use these data to regress the observed values of Y_{iM} on $\bar{L}_{i,M-1}$ and $\hat{\pi}_{iM}$ to obtain a consistently estimated model $\tilde{m}_M^{M-1}(\bar{L}_{i,M-1}, \hat{\pi}_{iM})$. For all subjects with $R_{i,M-1} = 1$ let

$$\tilde{Y}_{i,M}^{M-1} = \tilde{m}_M^{M-1}(\bar{L}_{i,M-1}, \hat{\pi}_{i,M}).$$

- 3 **For $s = M - 2, \dots, 1$ recursively:**

- 4 Identify all subjects i with $R_{i,s+1} = 1$, and regress the values of $\tilde{Y}_{i,M}^{M-1}$ on \bar{L}_{is} and $\hat{\pi}_{i,s+1}$ to obtain a consistently estimated model $\tilde{m}_{iM}^s(\bar{L}_{is}, \hat{\pi}_{i,s+1})$. For all subjects with $R_{is} = 1$, let

$$\tilde{Y}_{i,M}^s = \tilde{m}_{iM}^s(\bar{L}_{i,s}, \hat{\pi}_{i,s+1}).$$

- 5 **Final step:** Let $\hat{Y}_{iM}^{BR} = \tilde{Y}_{iM}^1$. Substitute \hat{Y}_{iM}^{BR} into the full data estimating functions U , and solve to obtain $\hat{\mu}_{iM}^{BR} = (1/N) \sum_{i=1}^N \hat{Y}_{iM}^{BR}$.
-

4.3. *Bang and Robins' doubly-robust recursive regression estimator for $E(Y_{iM})$ and extensions.* Bang and Robins (2005) introduced a regression-based approach for longitudinal data with dropout in the particular case where the estimand of interest is $E(Y_{iM})$. The estimator, given in Algorithm 2, uses recursive regression to impute the values of Y_{iM} , including $\hat{\pi}_{ij}^{-1}$ as a covariate in each imputation model in order to achieve double robustness. The algorithm differs from Paik's sequential mean imputation in that it uses imputed values \hat{Y}_{iM} as the outcome in each estimation step, rather than as predictors, and uses \hat{Y}_{iM} , even when observed values Y_{iM} are available. The estimator is shown to be asymptotically equivalent to the doubly-robust AIPW estimator from equation (7) in the cross-sectional setting (Bang and Robins (2005)). Although Bang and Robins (2005) only provided the case of $U_i = Y_{iM} - E(Y_{iM})$, they noted that this approach can be generalized by modifying the U_i and its corresponding score equation.

Schnitzer, Lok and Bosch (2016) extended the Bang and Robins approach to estimate the coefficients β , giving the association between Y_{iM} and baseline covariates X_{i0} of interest, in the context of a generalized linear model. The idea is to take $U_i = X_{i0}(Y_{iM} - g(\beta X_{i0}))$, with g an appropriate link function and where X_{i0} includes an intercept. The model for $\tilde{m}^s(\cdot)$ is also taken to use the same link function $g(\cdot)$. Then the approach is similar to Algorithm 2; however, rather than including $\hat{\pi}_{i,s+1}$ as a predictor in the model for \tilde{m}^s , one includes $\hat{\pi}_{i,s+1}^{-1} X_{i0}$. Finally, substitute \tilde{Y}_{iM}^1 into the estimating functions U_i , and solve to obtain the estimate of $\hat{\beta}$. For the special case $U_i(\bar{L}_{iM}) = Y_{iM} - E(Y_{iM})$, this recovers the Bang and Robins (2005) estimator.

We note here that, when applying the Bang and Robins approach for longitudinal data to estimate $E(Y_{iM})$, Tsiatis, Davidian and Cao (2011) used forward selection to select variables in both the logistic regression models for estimating λ_j and in the OLS models for estimating $E(Y_{iM} | \bar{L}_{ij})$. Similarly, in their implementation Schnitzer and colleagues performed a variable selection step in each regression. In our simulations we incorporate a similar variable selection step for these estimators, as it greatly improved their performance in practice. We denote this slightly modified and extended estimator **BR***. Also note that the estimator

requires the sequential regression models \tilde{m}^s in Algorithm 2 to be consistent with the estimating equation U that defines the estimand of interest; thus, a change of estimating equation may require a new set of sequential regressions. This seems to be a key difference from the imputation framework proposed below.

5. An imputation framework for longitudinal data using AIPW estimators. Here we show that the optimal AIPW estimating equations (7) can be written in a form that applies the full data estimating functions $U(\boldsymbol{\beta}, \bar{L}_{iM})$ to the “completed” data in which all levels of missing observations have been filled in using a set of doubly-robust imputed observations. In this approach $U(\boldsymbol{\beta}, \bar{L}_{iM})$ may be any equation of interest linear in \mathbf{Y}_i , including a GEE. The approach has the advantage that standard software can be used to solve for the estimates $\hat{\boldsymbol{\beta}}^{\text{AIPW}}$, and it provides a framework for flexible construction of AIPW estimators. In particular, unlike other doubly-robust approaches, when the parameters of interest change, there is no need to run the doubly-robust estimating procedure again.

5.1. The optimal AIPW estimating equations in imputation form. Consider the estimating functions $U(\boldsymbol{\beta}, \bar{L}_{iM}) = q(\boldsymbol{\beta})(\mathbf{Y}_i - \boldsymbol{\mu}_i)$, where $\mathbf{Y}_i = (y_{i,1}, y_{i,2}, \dots, y_{i,M})^T$ and $\boldsymbol{\mu}_i = (\mu_{i,1}, \mu_{i,2}, \dots, \mu_{i,M})^T$; for example, $q(\boldsymbol{\beta}) = \frac{\partial \mu_i^T}{\partial \boldsymbol{\beta}} V_i^{-1}$ with $\mu(\boldsymbol{\beta}) = g(\boldsymbol{\beta}X)$ when $U(\boldsymbol{\beta}, \bar{L}_{iM})$ is a GEE. Then one can show that the optimal AIPW estimator, given by equations (7) and (8), has an equivalent substitution form,

$$(10) \quad \sum_{i=1}^N U(\hat{\boldsymbol{\beta}}, \bar{L}_{i,M}^{\text{AIPW}}) = \sum_{i=1}^N q(\hat{\boldsymbol{\beta}})(\hat{\mathbf{Y}}_i^{\text{AIPW}} - \boldsymbol{\mu}_i) = 0,$$

where $\hat{\mathbf{Y}}_i^{\text{AIPW}}$ is a corresponding AIPW estimator of the full data \mathbf{Y}_i , given in (14) below. Specific examples of such AIPW imputation estimators are given in Section 6.

To demonstrate (10), first, it is straightforward to show that

$$(11) \quad \frac{C_{i,M}}{\pi_{i,M}} + \sum_{j=1}^{M-1} \left(\frac{C_{ij} - \lambda_{i,j+1} R_{ij}}{\pi_{i,j+1}} \right) = 1.$$

Thus, letting $\pi_{i,M+1} = \pi_{iM}$ and $\lambda_{i1} = \lambda_{i,M+1} = 0$, equation (7) can be rewritten as

$$(12) \quad \sum_{i=1}^N \sum_{j=1}^M \left(\frac{C_{ij} - \lambda_{i,j+1} R_{ij}}{\pi_{i,j+1}} \right) H^j(\boldsymbol{\beta}, \bar{L}_{ij}) = 0.$$

Next, following Seaman and Copas (2009) (for the case where $q(\boldsymbol{\beta}) = \frac{\partial \mu_i^T}{\partial \boldsymbol{\beta}} V_i^{-1}$), we have that (8) can be written as

$$H^j(\boldsymbol{\beta}, \bar{L}_{ij}) = q(\boldsymbol{\beta})(E[\mathbf{Y}_i | \bar{L}_{ij}] - \boldsymbol{\mu}_i).$$

Then substituting into (12), switching the order of summation, and recognizing that $\sum_{j=1}^M (C_{ij} - \hat{\lambda}_{j+1} R_{ij}) / \hat{\pi}_{j+1} = 1$, we obtain that (7) can be written as

$$(13) \quad q(\boldsymbol{\beta}) \left\{ \left(\sum_{j=1}^M \left(\frac{C_{ij} - \lambda_{i,j+1} R_{ij}}{\pi_{i,j+1}} \right) E[\mathbf{Y}_i | \bar{L}_{ij}, R_{ij} = 1] \right) - \boldsymbol{\mu}_i \right\}.$$

Finally, we can write

$$(14) \quad \hat{\mathbf{Y}}_i^{\text{AIPW}} = \sum_{j=1}^M \left(\frac{C_{ij} - \hat{\lambda}_{i,j+1} R_{ij}}{\hat{\pi}_{i,j+1}} \right) E[\hat{\mathbf{Y}}_i | \bar{L}_j, R_j = 1]$$

to obtain (10), recognizing \hat{Y}_i^{AIPW} as the representation of an optimal AIPW estimator of $\beta = E[\mathbf{Y}_i]$.

Form (10) has the advantage over the original approach in Seaman and Copas (2009) that, once the values of \hat{Y}_i^{AIPW} are obtained, standard software for the estimating functions $U(\beta, \bar{L}_{iM})$ can be used to solve for the doubly-robust estimator β^* . Furthermore, the form of \hat{Y}_i^{AIPW} is independent of the details of the estimating functions U , an independence that we can exploit to address additional estimands, defined by different estimating functions, using the same set of values \hat{Y}_i^{AIPW} . From the theory in Tsiatis (2006), we can be assured that β^* is a member of the class of locally efficient DR estimators. We may also use the above considerations to provide a direct demonstration that a general estimator $\hat{\beta}^{\text{AIPW}}$ is doubly-robust. A formal development of this proof can be found in Section 1 of the Supplementary Material (Qiu and Messer (2023)).

6. Imputation framework with two specific DR estimators. Here we apply the AIPW imputation framework to construct two particular DR estimators for longitudinal data with monotone dropout.

6.1. *AIPW-I: Sequential mean imputation.* We denote the sequential mean imputation implementation of the DR imputation estimator as $\hat{\beta}^{\text{AIPW-I}}$. The procedure can be described in two steps:

Imputation step: For each subject i , impute a doubly-robust complete data vector using the AIPW estimate of Y_{ik} , $k = 2, \dots, M$

$$(15) \quad \hat{Y}_{ik}^{\text{AIPW}} = \frac{C_{ik}Y_{ik}}{\hat{\pi}_{ik}} + \sum_{j=1}^{k-1} \left(\frac{C_{ij} - \hat{\lambda}_{ij+1}R_{ij}}{\hat{\pi}_{ij+1}} \right) \hat{m}_k^j(\bar{L}_{ij})$$

with the models $\hat{m}_k^j(\cdot) = E[Y_{ik} | \bar{L}_{ij}, R_{ij} = 1]$ estimated by Paik’s sequential mean regression as in Algorithm 1.

Estimation step: Substitute \hat{Y}_i^{AIPW} into the full data estimating equations. These equations are then solved using standard software to obtain the doubly-robust estimator $\hat{\beta}^{\text{AIPW-I}}$.

6.2. *AIPW-S: A computationally simpler baseline \times time imputation model.* Our second implementation is computationally simpler, denoted as $\hat{\beta}^{\text{AIPW-S}}$. It uses only baseline covariates and time for the imputation models, at the cost of potential efficiency loss and stronger required assumptions. This is motivated by the clinical trials setting in which it may be desirable to restrict attention to modeling with prerandomization covariates. For the imputation models, it uses up to $M - 1$ models with baseline covariates and time rather than the $M(M - 1)/2$ sequential recursive regression models required by $\hat{\beta}^{\text{AIPW-I}}$ or the Bang and Robins estimator. It can be thought of as a simplified version of the fully efficient approach, given in Tsiatis, Davidian and Cao (2011). If the covariates are sufficient to render the data MAR, then $\hat{\beta}^{\text{AIPW-S}}$ will be doubly robust. The cost is potential efficiency loss and a stronger assumption regarding the MAR conditions. In the simulations and application, we give an example of the simplest implementation, in which only one single MMRM is estimated rather than $M - 1$ models or $M(M - 1)/2$ models, as in the AIPW-I and related estimators.

Here we give a sketch of the development; please refer to Section 2 of the Supplementary Material (Qiu and Messer (2023)) for details. To develop the estimator, we start with (13) and

note that $\hat{E}[Y_{ik}|\bar{L}_{ij}] = Y_{ik}$ for $j \geq k$. Hence, we can write the k th component of $\hat{\mathbf{Y}}_i^{\text{AIPW}}$ as

$$(16) \quad \hat{Y}_{ik}^{\text{AIPW}} = \sum_{j=k}^M \left(\frac{C_{ij} - \hat{\lambda}_{i,j+1} R_{ij}}{\hat{\pi}_{i,j+1}} \right) Y_{ik} + \sum_{j=1}^{k-1} \left(\frac{C_{ij} - \hat{\lambda}_{i,j+1} R_{it}}{\hat{\pi}_{i,j+1}} \right) \hat{Y}_{ik}^j.$$

Next, we take \hat{Y}_{ik}^j to be a consistent estimate of $E[Y_{ik}|X_{i0}, t]$ independent of j , where X_{i0} contains baseline covariates and t indicates time. This amounts to the choice $H^j(\boldsymbol{\beta}, \bar{L}_{ij}) = E[Y_{ik}|X_{i0}, t]$. Note that, while we are now outside the set of possible efficient estimators, given by equation (7), except in the special case where $E[Y_{ik}|\bar{L}_{ij}] = E[Y_{ik}|X_{i0}, t]$, the arguments for double robustness remain unchanged. (Please see Section 2 of the Supplementary Material (Qiu and Messer (2023)) for details.)

In this case (16) simplifies to a repeated cross-sectional form,

$$(17) \quad \hat{Y}_{ik}^{\text{AIPW-S}} = \frac{R_{ik}}{\hat{\pi}_{ik}} Y_{ik} + \left(1 - \frac{R_{ik}}{\hat{\pi}_{ik}} \right) \hat{Y}_{ik},$$

because $\sum_{j=k}^M ((C_{ij} - \hat{\lambda}_{i,j+1} R_{ij})/\hat{\pi}_{i,j+1}) = R_{ik}/\hat{\pi}_{ik}$.

The procedure can be described in two steps:

Imputation step: For each subject i , impute a doubly-robust complete data vector using the AIPW-S estimate of Y_{ik} , $k = 2, \dots, M$, given by (17), with $\hat{Y}_{ik} = \hat{m}_k(X_{i0}, t)$, depending only on baseline covariates X_{i0} and time t , and where $m_k(X_{i0}, t) = E[Y_{ik}|X_{i0}, t]$ is estimated from the observed data.

Estimation step: Substitute $\hat{\mathbf{Y}}_i^{\text{AIPW-S}}$ into the full data estimating equations. These equations are then solved using standard software to obtain the doubly-robust estimator $\hat{\boldsymbol{\beta}}^{\text{AIPW-S}}$.

A detailed development can be found in Section 2 of the Supplementary Material (Qiu and Messer (2023)). In the simulation and applications below, we give examples where a single mixed-effects model $\hat{m}(X_0, t)$ is estimated using all observed responses as the outcomes, regressed on baseline covariates X_0 and time t . Then $\hat{Y}_{ik} = \hat{m}(X_{i0}, t)$, and formula (17) is used to derive a completed doubly-robust data set. In this case, instead of $M(M-1)/2$ regressions as for the other sequential regression estimators, only one regression needs to be estimated.

7. Simulations. We use simulation to investigate the performance of the two DR imputation-framework estimators in the setting of normal longitudinal generalized estimating equations under MAR monotone dropout: **AIPW-I**, based on AIPW using Paik's sequential mean imputation (Section 6.1), and **AIPW-S**, AIPW with a computationally simpler imputation model using only baseline covariates (Section 6.2). We also compare the use of the Bang and Robins regression-based estimator **BR*** in both its original form and as extended by Schnitzer, Lok and Bosch (2016) to more general estimating equations for generalized linear models.

We study two different estimands in two sets of simulations: first, $E[Y_k]$ for the case $k = 1, \dots, M$ in which case **BR*** reduces to the original Bang and Robins estimator, and second, the vector of regression coefficients $\boldsymbol{\beta}$ defined as the solution (using the full-data distribution) to the score equations U , given by a mixed model with repeated measures (MMRM), similar to what might be used in the primary analysis of a clinical trial. In particular, $\boldsymbol{\beta}$ includes a time by treatment interaction which represents an important estimand in the clinical trials setting.

For comparison, we also include two traditional observed-data regression estimators often used in clinical trials: (1) a mixed model of repeated measures (**MMRM**) with random

intercept and slope and (2) a generalized estimating equations model using an independence working correlation (**GEE-IND**) with a sandwich estimator of variance. As negative controls, which adjust for dropout but are not DR, we also include Paik's sequential mean imputation (**Paik**), as described in Section 3.4, and inverse probability weighted GEE (**WGEE**), as implemented in the R package `wgeesel` (Xu et al. (2019)). Finally, as a further negative control in the imputation framework, we use data imputed using **BR*** in the first simulation for estimating a time by treatment interaction in the second simulation. As **BR*** is not designed to work as an imputation estimator, under model misspecification this may demonstrate the utility of the new estimators, in practice, in a special case. Notably, we allow the imputation model to differ from the estimation model in order to show the utility of the imputation framework.

The two sets of simulations are linked for the imputation estimators. We used the estimates of $E[Y_k]$ from the first set of simulations to obtain a completed dataset. We then used a standard GEE model with the completed data to estimate the regression coefficients β in the second set of simulations.

For each simulated data set and estimation method, we compute both a point estimate for the estimand of interest and an associated 95% normal-theory confidence interval, where the variance of the parameter is computed using the nonparametric bootstrap. We consider both a moderate dropout construct ($\approx 30\%$ dropout) and an extreme dropout construct ($\approx 50\%$ dropout). To investigate double robustness, we construct four scenarios, depending on whether the dropout models and/or the imputation models are specified correctly or incorrectly.

Importantly, we include in the study an extreme form of model misspecification, as described in 7.2 and 7.3, in which the incorrect dropout models or imputation models do not include the treatment indicator, and the estimand is the coefficient of a treatment by time interaction. This case allows the estimand analysis model to differ substantially from the imputation model and illustrates what might happen when using a preexisting imputed complete dataset to study a different estimand of interest.

7.1. Performance metrics and sample sizes. We report Monte Carlo estimates (500 repeats) for the bias, standard deviation, and root mean square error (RMSE) of the point estimate. For the confidence interval, we report the coverage probability, the mean of the bootstrap standard error estimate, and the mean interval score of [Gneiting and Raftery \(2007\)](#), given by

$$(18) \quad S(\hat{l}, \hat{u}, \theta) = (\hat{u} - \hat{l}) + \frac{2}{\alpha} ((\hat{l} - \theta)\mathbb{1}\{\theta < \hat{l}\} + (\theta - \hat{u})\mathbb{1}\{\hat{u} < \theta\}),$$

where θ is the true parameter of interest, $1 - \alpha$ is the nominal confidence level, (\hat{l}, \hat{u}) are the interval limits, and $\mathbb{1}\{\}$ denotes the indicator function. A MMRM model with correct covariates and random intercept and slope will serve as the gold standard in our comparisons; this reflects the complete data generating model.

7.2. Specification of the data generating model, the primary estimand, and the correct and incorrect imputation models. Longitudinal responses Y_{ij} are generated from a mixed-effects model, using similar parameters as in [Tsiatis, Davidian and Cao \(2011\)](#), as

$$(19) \quad Y_{ij} = b_{0i} + b_{1i}t + \beta_0 + \beta_1x_1 + \beta_2x_2 + \beta_3x_2 \times t + \epsilon.$$

The covariates are generated as $x_1 \sim N(5,1)$, $x_2 \sim \text{Bernoulli}(0.5)$, with $\epsilon \sim N(0, 1)$. Considering the clinical trial setting, x_2 would indicate the treatment variable and x_1 a continuous covariate. The sample size is $n = 500$, and three time points are $t = 1, 2, 3$. The bootstrap

sample size (for variance estimation) is 300. Here b_0 and b_1 are random intercepts and slopes from a bivariate normal with mean $\mu = \begin{pmatrix} 1 \\ 6 \end{pmatrix}$ and covariance $\Sigma = \begin{pmatrix} 0.3 & 0.1 \\ 0.1 & 0.2 \end{pmatrix}$. The data generating coefficients are $\beta_0 = 0.5$, $\beta_1 = 2$, $\beta_2 = -0.25$, $\beta_3 = -6$. Thus, the expectation of Y_{ij} is 11.375, 14.375, 17.375 for $j = 1, 2, 3$.

Only a few studies have performed longitudinal simulations to evaluate AIPW estimators. Comparing with Tsiatis, Davidian and Cao (2011) and others, we further include an arm by time interaction, providing a more realistic model for randomized trials. The primary estimand is either $E[Y_3]$ or the vector of regression coefficients $\beta_{\text{est}} = (E(b_{1i}), \beta_2, \beta_3)$, obtained from the estimating equations for model (19), referring to the slope of time, difference between arms at baseline, and effect of treatment along with time.

For all methods the correctly specified mean model includes categorical time (if needed), (x_1, x_2) , and the interaction of time by x_2 . For sequential imputation methods, the correct models are of the form $Y_j \sim Y_1 + \dots + Y_{j-1} + x_1 + x_2 + e$ for $j = (2, 3)$. For all methods the misspecified imputation model excludes the treatment indicator x_2 and the corresponding interaction terms. Note that, in our setting, this is a case of severe model misspecification, as the difference between treatment arms is of primary interest. From another perspective there is a severe mismatch between the imputation models m , which omit the variable of primary interest, and the estimating equations U which define the estimand. Thus, this may also be thought of as a test of the robustness of the imputation framework to changing the estimand of interest after the imputation step has been carried out. Details of the model specifications can be found in Section 3 of the Supplementary Material (Qiu and Messer (2023)).

7.3. Specification of the dropout generating model and of correct and incorrect models for missingness. We generate dropout according to the logistic regression models

$$\text{logit}(\lambda_2) = \gamma_{20} + \gamma_{21}y_1 - \gamma_{22}x_2 + e,$$

$$\text{logit}(\lambda_3) = \gamma_{30} + \gamma_{31}y_1 + \gamma_{32}y_2 - \gamma_{33}x_2 + e,$$

where λ_{ij} is the hazard function of dropout, as defined in formula (3). For the moderate dropout construct, $(\gamma_{20}, \gamma_{21}, \gamma_{22}, \gamma_{30}, \gamma_{31}, \gamma_{32}, \gamma_{33}) = (-7.625, 0.5, 2, -5.225, 0.1, 0.2, 4)$. The empirical mean dropout rates from 500 Monte Carlo repeats were 10% for Y_2 and 30% for Y_3 . For the high dropout construct, $(\gamma_{20}, \gamma_{21}, \gamma_{22}, \gamma_{30}, \gamma_{31}, \gamma_{32}, \gamma_{33}) = (-7, 0.5, 1, -4.5, 0.1, 0.2, 2)$ and the mean dropout rates were 20% and 48% for Y_2 and Y_3 , respectively. Misspecified dropout models omit the treatment indicator x_2 which again omits the difference between treatment arms in the clinical trials setting.

7.4. Specification of the BR estimator used as an imputation estimator.* In the case when the estimand of interest is the coefficient of a baseline covariate X and where the imputation models are consistent with the estimating equations, the BR* estimator, given in Section (4.3), is similar in form to an imputation framework estimator. Hence, we investigated empirically the use of BR* in the imputation step (15). This can be done by setting M in Algorithm 2 sequentially from $M = 2$ to the final visit and then substituting each $\hat{Y}_{ik}^{\text{AIPW}}$ in (15) by $\hat{\mu}_{ik}^{\text{BR}}$. Note that, in the case of estimating the coefficient of a treatment by time interaction term, we are outside of the category of estimands considered by Bang and Robins (2005) and Schnitzer, Lok and Bosch (2016). Also, the imputation framework allows the model in the imputation step to differ from the estimating equations model and so in this case is again outside the class of estimators considered in those papers. In our simulations we used forward variable selection in the imputation models, following the implementations in both Tsiatis, Davidian and Cao (2011) and Schnitzer, Lok and Bosch (2016), as this, in practice, improved performance.

TABLE 1

Comparison by simulation of point and interval estimators for $E(Y_3)$ under a moderate dropout rate of 30% missing. Performance metrics are: Bias, Root mean square error (RMSE), the interval score (IntS), coverage probability (CovP), Monte Carlo standard deviation (MCSD), and mean bootstrap estimate of standard error (AveSE) from 500 simulation runs. The sample size was $n=500$, and the bootstrap sample size was 300. The true value of $E(Y_3) = 17.4$

	Bias	RMSE	IntS	CovP	MCSD	AveSE	Bias	RMSE	IntS	CovP	MCSD	AveSE
	<i>Y correct P correct</i>						<i>Y correct P incorrect</i>					
BR*	-0.01	0.30	1.37	0.95	0.30	0.30	-0.01	0.30	1.37	0.95	0.30	0.30
AIPW-I	-0.01	0.30	1.39	0.95	0.30	0.31	-0.00	0.30	1.37	0.95	0.30	0.30
AIPW-S	-0.01	0.31	1.39	0.95	0.31	0.31	0.04	0.31	1.39	0.95	0.31	0.31
Paik	-0.01	0.30	1.35	0.95	0.30	0.30	-0.01	0.30	1.35	0.95	0.30	0.30
MMRM	-0.01	0.30	1.38	0.95	0.30	0.30	-0.01	0.30	1.38	0.95	0.30	0.30
WGEE	0.00	0.30	1.37	0.95	0.30	0.30	0.03	0.31	1.39	0.95	0.31	0.31
GEE-IND	-0.09	0.31	1.40	0.94	0.29	0.30	-0.09	0.31	1.40	0.94	0.29	0.30
	<i>Y incorrect P correct</i>						<i>Y incorrect P incorrect</i>					
BR*	-0.00	0.33	1.54	0.95	0.33	0.34	-0.64	0.71	6.06	0.53	0.31	0.33
AIPW-I	-0.01	0.31	1.44	0.95	0.31	0.31	-0.68	0.74	7.65	0.42	0.31	0.31
AIPW-S	-0.04	0.38	1.72	0.95	0.38	0.36	-0.62	0.71	6.44	0.55	0.36	0.36
Paik	-0.65	0.72	6.95	0.45	0.31	0.32	-0.65	0.72	6.95	0.45	0.31	0.32
MMRM	-0.60	0.67	5.83	0.50	0.31	0.32	-0.60	0.67	5.83	0.50	0.31	0.32
WGEE	0.00	0.32	1.44	0.94	0.32	0.32	-0.62	0.71	6.11	0.53	0.34	0.36
GEE-IND	-2.18	2.20	63.84	0.00	0.30	0.31	-2.18	2.20	63.94	0.00	0.30	0.31

7.4.1. Results for estimating $E(Y_3)$ under moderate dropout. The upper-left panel of Table 1 shows results for estimating $E(Y_3)$ under the moderate dropout construct when both imputation and dropout models are correct. The “gold standard” MMRM model (correct complete data maximum likelihood estimator with MAR data) has a bias of -0.01 . GEE-IND has a worse bias of -0.09 , which supports the theory that, even if the mean structure is correct, an incorrect working correlation may still cause bias under MAR longitudinal dropout. The bias of all three doubly-robust methods is less than 0.01, and the performance is very similar to the gold standard, consistent with their asymptotic local efficiency. Paik’s imputation and WGEE also perform well in terms of bias and efficiency. However, across all moderate dropout scenarios, approximately 1% of Monte Carlo repeats for WGEE reported convergence issues, sometimes resulting in substantial standard errors and outlier estimates. The program also consumes much more time than other methods, by a factor of about 5.

When the imputation model is correct but the dropout model is misspecified (upper-right panel), all doubly-robust methods have acceptable bias, ranging from 0.037 to -0.005 , with RMSEs, coverage probabilities, interval scores, Monte Carlo standard deviations, and average estimated standard errors all similar to the gold standard, indicating the estimators are consistent and remain close to efficient.

When the dropout model is correct but the imputation model is misspecified (bottom-left panel), all doubly-robust methods again have acceptable bias, indicating their double robustness. AIPW-I has the best efficiency, and AIPW-S loses some efficiency, compared to prior scenarios. In this scenario the original BR estimator without variable selection had an unacceptably large bias and low coverage probability (85%, not shown), although the modified estimator BR* (Tsiatis, Davidian and Cao (2011)) performs well, as shown in the Table 1. As expected, the nondoubly-robust regression-based methods did not work well. The incorrect MMRM model and Paik’s imputation have coverage probabilities less than 50%, with large bias and bad efficiencies. A GEE model with wrong mean structure and wrong working correlation matrix performed the worst.

TABLE 2

Comparison by simulation of point and interval estimators for regression coefficients β under a moderate dropout rate of 30% missing. Performance metrics are: Bias, Root mean square error (RMSE), the interval score (IntS), and coverage probability (CovP) for a 95% confidence interval from 500 simulation runs. The sample size was $n = 500$, and the bootstrap sample size was 300. The true values of the parameters β are 2, -0.25, and 6 for the coefficients of x_2 (treatment indicator), time, and the x_2 :time interaction term, respectively

	Coefficient of x_2				Coefficient of time				Coefficient of time $\times x_2$			
	Bias	RMSE	IntS	CovP	Bias	RMSE	IntS	CovP	Bias	RMSE	IntS	CovP
	Y correct P correct											
BR*	0.01	0.13	2.24	0.98	0.00	0.11	0.50	0.93	-0.01	0.14	0.67	0.94
AIPW-I	0.01	0.10	0.50	0.95	0.00	0.11	0.50	0.94	-0.01	0.14	0.68	0.93
AIPW-S	0.01	0.10	0.50	0.95	0.00	0.11	0.50	0.95	-0.01	0.14	0.68	0.93
Paik	0.01	0.10	0.50	0.95	0.00	0.10	0.50	0.93	-0.01	0.14	0.67	0.93
MMRM	0.01	0.10	0.50	0.94	0.00	0.10	0.50	0.94	-0.01	0.14	0.66	0.94
WGEE	0.01	0.10	0.50	0.95	-0.01	0.10	0.50	0.95	0.00	0.14	0.65	0.94
GEE-IND	0.01	0.10	0.50	0.95	-0.03	0.11	0.51	0.94	0.02	0.14	0.64	0.95
	Y correct P incorrect											
BR*	0.01	0.18	1.70	0.98	0.00	0.10	0.50	0.94	-0.01	0.14	0.66	0.94
AIPW-I	0.01	0.10	0.50	0.95	0.00	0.10	0.50	0.94	-0.01	0.14	0.68	0.93
AIPW-S	0.01	0.10	0.50	0.95	-0.01	0.10	0.50	0.94	0.02	0.14	0.66	0.94
Paik	0.01	0.10	0.50	0.95	0.00	0.10	0.50	0.93	-0.01	0.14	0.67	0.93
MMRM	0.01	0.10	0.50	0.94	0.00	0.10	0.50	0.94	-0.01	0.14	0.66	0.94
WGEE	0.01	0.10	0.51	0.95	-0.01	0.10	0.50	0.95	0.02	0.14	0.65	0.95
GEE-IND	0.01	0.10	0.50	0.95	-0.03	0.11	0.51	0.94	0.02	0.14	0.64	0.95
	Y incorrect P correct											
BR*	0.00	0.14	2.79	0.98	-0.83	0.89	20.82	0.04	1.65	1.75	61.60	0.00
AIPW-I	0.01	0.10	0.50	0.95	0.00	0.11	0.53	0.95	-0.01	0.15	0.70	0.94
AIPW-S	0.01	0.11	0.51	0.95	0.00	0.11	0.55	0.95	-0.01	0.15	0.71	0.94
Paik	0.01	0.10	0.51	0.95	-0.59	0.61	13.43	0.01	0.67	0.69	15.12	0.01
MMRM	0.25	0.25	10.00	0.00	-3.24	3.25	117.84	0.00	6.00	6.00	240.00	0.00
WGEE	0.25	0.25	10.00	0.00	-3.07	3.08	111.33	0.00	6.00	6.00	240.00	0.00
GEE-IND	0.25	0.25	10.00	0.00	-3.30	3.31	120.29	0.00	6.00	6.00	240.00	0.00
	Y incorrect P incorrect											
BR*	0.02	0.25	3.59	0.97	-3.30	3.31	119.88	0.00	6.08	6.08	238.57	0.00
AIPW-I	0.01	0.11	0.50	0.95	-0.29	0.31	3.38	0.34	0.05	0.15	0.67	0.95
AIPW-S	0.01	0.11	0.50	0.95	-0.32	0.34	4.05	0.28	0.10	0.18	0.79	0.91
Paik	0.01	0.10	0.51	0.95	-0.59	0.61	13.43	0.01	0.67	0.69	15.12	0.01
MMRM	0.25	0.25	10.00	0.00	-3.24	3.25	117.84	0.00	6.00	6.00	240.00	0.00
WGEE	0.25	0.25	9.98	0.00	-3.27	3.27	118.58	0.00	6.00	6.00	239.97	0.00
GEE-IND	0.25	0.25	10.00	0.00	-3.30	3.31	120.29	0.00	6.00	6.00	240.00	0.00

The bottom-right panel shows results when both models are incorrect. All methods have worse performance than in other scenarios, indicating that the misspecification in the models is substantial and thus provides a good test of double robustness. In this scenario all methods report similar bias and efficiencies, except GEE-IND, which reports a much worse result.

7.4.2. *Results for estimating coefficients β under moderate dropout.* Table 2 presents results for estimating the vector of three regression coefficients $\beta = (E(b_1), \beta_2, \beta_3)$, as defined in (19). The three regression coefficients are the coefficients of time, x_2 , and x_2 by time. In a randomized trial, these coefficients would correspond to the time trend for the placebo arm, the baseline difference between arms, and the treatment effect on the time trend.

With both models correctly specified, the estimated $\hat{\beta}$ appears to be consistent and efficient in all methods, except that the Bang and Robins–Schnitzer approach (BR*) reports a larger

variance for the estimated β_2 , the coefficient of the treatment indicator x_2 . After excluding simulation runs with substantial standard errors (greater than 50), BR* still has a large mean interval score and some unusual standard errors from Monte Carlo repeats.

When the imputation model is correctly specified but the dropout model incorrectly excludes the treatment indicator x_2 , the performance of the three doubly-robust estimators is as good as the first scenario, indicating that all three doubly-robust imputed data sets have minimal bias in this scenario. While AIPW-I and AIPW-S also obtain comparable efficiencies to the gold standard, BR* again has larger RMSE and mean interval scores when estimating β_2 , the treatment arm indicator.

An interesting case occurs when the dropout model is correctly specified but the imputation model omits x_2 and the time \times x_2 interaction. First, consider the coefficient of the baseline indicator x_2 , where all three estimators AIPW-I, AIPW-S, and BR* perform as before, with minimal bias in accordance with their double-robustness. Next, consider the coefficients of the omitted terms time and time \times x_2 . Here AIPW-I and AIPW-S continue to obtain unbiased and efficient estimators for the coefficients of these terms. This illustrates their double robustness and that the imputation model does not need to accord with the estimation model. However, in this setting BR*, when used as an imputation estimator, fails to consistently estimate the coefficients for time and time \times x_2 . This is not entirely unexpected, because the BR* approach is not developed for time-varying predictors of interest and also because it requires the variables of interest in the final estimating equations to be included in the imputation models. Thus, the BR* estimator is indeed not robust to this particular kind of model misspecification. Paik's mean imputation derives an unbiased estimator for β_2 but fails to consistently estimate β_3 and $E(b_1)$. MMRM, WGEE, and GEE-IND do not perform well in this scenario.

When both models are misspecified, none of the methods obtain consistent estimates for β . As before, this indicates that the misspecification in the imputation and weighting models is substantial and thus provides a good test of double robustness. Notably, in our simulation AIPW-I and AIPW-S are more robust than other methods, especially for estimating $\hat{\beta}$ for x_2 and time by x_2 , the treatment and treatment by time interaction terms.

7.5. Extreme dropout construct. We conduct a similar comparison of these estimators in the extreme dropout construct, where the mean dropout rates for Y_2 is 20% and for Y_3 is 48%. Other simulation parameters are kept as previously described. Results are qualitatively similar to the moderate dropout construct with generally good performance of the three doubly-robust estimators, compared to the MMRM estimator when estimating $E[Y_3]$, and good performance of AIPW-I and AIPW-S when estimating β through the fully imputed data sets. In some scenarios the loss of efficiency of AIPW-S relative to AIPW-I becomes apparent, although its performance is still acceptable. Compared with the moderate construct, WGEE has even more convergence issues reported. Details are given in Section 4 of the Supplementary Material (Qiu and Messer (2023)).

8. Application to the MCI trial. The MCI trial has been described in Section 2. The primary outcome of the trial was time to progression to Alzheimer's disease (AD). The main conclusion of the trial was that Vitamin E had no benefit, while donepezil provided some benefit over placebo at 12 months but not at 36 months, in accordance with the known symptomatic benefits of donepezil. The trial showed no benefit in secondary analyses comparing within-patient change on the two cognitive measures Mini-Mental State Examination (MMSE) and the Clinical Dementia Rating sum of boxes (CDR-SOB) at 36 months.

Here, for simplicity, only the donepezil arm and placebo arm are studied. The two arms have similar demographic and clinical characteristics at baseline. The within-subject change

from baseline for the MMSE and the CDR-SOB are used as our repeated measures outcomes. Lower MMSE (range: 0 to 30) and higher CDR-SOB (range: 0 to 18) indicate worse cognition. These measures are assessed at baseline, 12 months, 24 months, and 36 months. The mean values of MMSE for placebo group vs. donepezil group are 27.3 (SD = 1.80) vs. 27.2 (SD = 1.86) at baseline, 26.6 (SD = 2.81) vs. 27.1 (SD = 2.46) at 12 months, 26.2 (SD = 3.48) vs. 26.4 (SD = 3.14) at 24 months, and 25.0 (SD = 5.05) vs. 25.3 (SD = 4.79) at 36 months. The mean values of CDR-SOB for placebo group vs. donepezil group are 1.87 (SD = 0.79) vs. 1.78 (SD = 0.80) at baseline, 2.28 (SD = 1.45) vs. 1.98 (SD = 1.22) at 12 months, 2.72 (SD = 2.02) vs. 2.56 (SD = 2.04) at 24 months, and 3.32 (SD = 3.09) vs. 3.20 (SD = 2.66) at 36 months. The missing rates for the donepezil arm and the placebo arm are 26.1% and 16.6% at 12 months, 34.0% and 29.3% at 24 months, and 42.7% and 32.0% at 36 months. In order to ensure that the dropout is monotone, we use Paik's mean imputation (Paik (1997)) to fill in the intermediate missing responses. Only about 1.5% of missing responses are intermediate missing values.

We compare results of the two doubly-robust imputation estimators, AIPW-I and AIPW-S, and an estimator from a GEE model with an independence working correlation (GEE-IND). We also present unadjusted estimates of the mean at each time point, using all observed data. All models include the baseline outcome score and model time as a linear trend, including both arm and the arm by time interaction as covariates, similar to one type of standard analysis model for AD trials. Two primary estimands of interest are considered for each outcome. The first estimand is the mean difference between arms, placebo arm–treatment arm, in within-patient change at the final visit (36 months) which is the primary estimand in many late stage AD trials. The least-squares mean difference, a model adjusted estimate, would be used (SAS Institute Inc. (2012)). The second estimand of interest is the regression coefficient for the interaction between treatment arm and time. In addition, the model-adjusted estimates of the average difference between arms at intermediate time points (12 and 24 months) and the coefficient of time are also presented, as these statistics would normally be computed in the final analysis of a clinical trial. The dropout model for the AIPW methods include arm, baseline outcome score, and historical Y as covariates. Normal theory confidence intervals are constructed using the nonparametric bootstrap estimates of variance; the bootstrap sample size is 500.

Figure 2 displays the least-squares mean estimates for the mean difference between arms in within-patient change for the MMSE (left panel) and the CDR-SOB (right panel) at 12, 24, and 36 months. The gray solid line shows the underlying mean differences at each time point in the observed data. For the MMSE comparison, a positive difference is in favor of active treatment; for the CDR-SOB comparison, a negative difference is in favor of active treatment. Consistent with the known transient benefit of donepezil, the raw data for both measures show some benefit to the active arm at the 12 month assessment which then diminishes toward zero at the 24 month and 36 month assessments.

The two doubly-robust imputation estimators (dotted and dashed) give similar results to each other, showing a modest average benefit from donepezil at the 12 months time point, as expected. In all cases this estimated early benefit tends to diminish over time, linearly because of the form of the model. The estimates made using the GEE-IND approach (black solid line) diminish faster than the doubly-robust estimates, resulting in a smaller estimated treatment effect at the last visit. In fact, the GEE-IND estimate at 36 months concluded that the CDR-SOB change between the donepezil arm and the placebo arm was 0.001, which is in nominal favor of the placebo arm, while the AIPW-I and AIPW-S imputation methods obtained estimated values of -0.078 and -0.059 in support of the donepezil arm. Table 3 presents the estimates and standard errors. Since all the covariates are baseline characteristics, the AIPW-S method shows good relative efficiency, with the smallest standard errors of 0.458 and 0.262 for MMSE and CDR-SOB at 36 months, respectively.

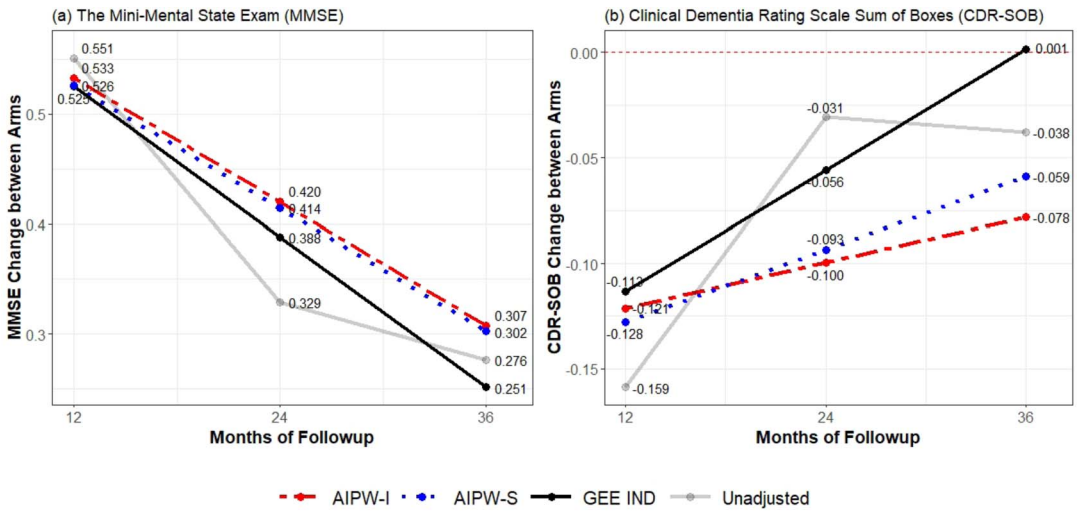


FIG. 2. Least-Squares Estimates of Mean Difference between Donepezil and Placebo Arms in Within-Patient Change, for the MCI trial. Left-hand panel shows MMSE; a positive difference is in favor of active treatment. Right-hand panel shows CDR-SOB; a negative difference is in favor of active treatment. In both cases the apparent early benefit of donepezil (12 months time point) appears to diminish over time. Gray solid line: Unadjusted mean difference, using the available data; Black solid line: Adjusted estimates from a GEE model with independent working correlation; Dotted line: Doubly-robust estimate using the simplified AIPW-S imputation for missing outcomes; Dashed line: Doubly-robust estimate using AIPW-I to adjust for missing outcomes.

Table 3 also shows the estimated values of β_{time} and $\beta_{time:arm}$ by the two AIPW imputation methods and the GEE-IND model. Again, the AIPW-I and AIPW-S obtained similar estimated values, while overall AIPW-S had smaller estimated standard errors. GEE-IND reported different estimated values and greater standard errors comparing with the two doubly-robust imputation methods.

9. Discussion. In this paper, using the approach of Seaman and Copas (2009), we have developed an imputation framework for AIPW-based doubly-robust estimators which is suitable for a class of general longitudinal estimating equations when the data are observed with monotone dropout under MAR. Confidence intervals are constructed using the bootstrap, and

TABLE 3

MCI trial, Estimates of Regression Coefficients for time and time:arm, and Least-Squares Estimates of Mean Difference between Arms in Within-Patient Change, at 12, 24, and 36 Months

	time		time:arm		Diff. in one-yr change		Diff. in two-yr change		Diff. in three-yr change	
	Estimate	SE	Estimate	SE	Estimate	SE	Estimate	SE	Estimate	SE
<i>Mini-Mental State Exam (MMSE)</i>										
GEE-IND	-0.803	0.148	-0.137	0.226	0.525	0.224	0.388	0.285	0.251	0.463
AIPW-I	-0.924	0.138	-0.113	0.224	0.533	0.223	0.420	0.290	0.307	0.468
AIPW-S	-0.909	0.138	-0.112	0.218	0.526	0.222	0.414	0.286	0.302	0.458
<i>Clinical Dementia Rating Scale (CDR) Sum of Boxes</i>										
GEE-IND	0.538	0.087	0.057	0.122	-0.113	0.113	-0.056	0.173	0.001	0.277
AIPW-I	0.632	0.081	0.022	0.111	-0.121	0.110	-0.100	0.168	-0.078	0.263
AIPW-S	0.625	0.081	0.034	0.110	-0.128	0.109	-0.093	0.168	-0.059	0.262

Standard errors were derived from 500 times nonparametric bootstrap.

the estimators can be implemented using standard software tools. We develop two specific imputation estimators, AIPW-I, and AIPW-S that appear to have reliable performance within the framework. The simpler AIPW-S estimator uses only baseline values in the imputation models and reduces the number of models that need to be estimated from $M(M - 1)/2$ to maximum $M - 1$, at the cost of some stronger assumptions and potential loss of efficiency. We show that the imputed completed data from either AIPW-I or AIPW-S can be substituted into a full data estimating equation $\sum_{i=1}^N U_i(\boldsymbol{\beta}) = 0$ to obtain an estimator which inherits doubly-robust properties from the imputation step. In particular, as is the case in other imputation settings, a given AIPW-imputed data set can be used across several different estimands, as the imputation models and the analysis models do not need to be coupled to one another. For example, a single AIPW-imputed data set might be used across several primary or secondary analyses in a clinical trial. Under a prespecified estimand, the AIPW-I estimator has the same workflow as the estimator of Seaman and Copas (2009); however, once a completed data set has been created, it can be used for variety of other estimands without refitting the doubly-robust model again. AIPW-S has considerably simplified computational demands, compared to AIPW-I.

We conducted extensive simulation studies across different estimands to evaluate the performance of these AIPW-based imputation estimators, including comparison with the approach proposed by Bang and Robins (BR*), which incorporates the probability weights inside the estimating equations as an additional covariate. The BR* estimator, as originally proposed, is suitable for estimating the outcome at a final time point, $E[Y_{iM}]$. Interestingly, we found only a few prior studies which make similar comparisons in the longitudinal setting (Tsiatis, Davidian and Cao (2011), Seaman and Copas (2009), Schnitzer, Lok and Bosch (2016)).

When the estimand of interest was $E[Y_{iM}]$ or was the regression coefficient of a baseline covariate, the three doubly-robust estimators, BR*, AIPW-I, and AIPW-S were consistent as expected, even when the imputation models or the dropout models were badly misspecified. Furthermore, these doubly-robust estimators appeared to be competitive in efficiency with the correctly specified MLE estimators, when the imputation models were correctly specified. Bootstrap confidence intervals had correct coverage probabilities with reasonable bootstrap sample sizes.

When the primary estimand was the coefficient of a time-varying covariate, the two AIPW imputation methods, AIPW-I and AIPW-S, again demonstrated performance comparable to efficient MLE estimators whenever the imputation models were correctly specified. When the imputation model was incorrect but the dropout model was correct, the two AIPW imputation estimators still performed well, with low bias and good coverage probabilities, although with some loss of efficiency. In particular, we studied an extreme case of a misspecified imputation model which omitted the time-varying covariate of interest (i.e., the estimand in the analysis model was the coefficient of a time-variable which was omitted from the imputation model). This illustrates a case when the imputation model might be constructed prior to considering the estimand of interest, such as could be the case when using already imputed data to study a new estimand. The two imputation-framework estimators worked well in this case.

In this last scenario, we also compared use of the BR* estimator for the imputation step within the imputation framework. Our rationale for doing this is because, in the case when the imputation model is correct and agrees with the analysis model, the BR* estimator is similar in form to an imputation estimator. However, when used as an imputation estimator in the extreme case described above, where the estimand in the analysis model was the coefficient of a time-varying variable which was omitted from the imputation model, the result had unacceptably large bias, compared to AIPW-I or AIPW-S. This is not unexpected, because the BR* estimator is, as yet, only demonstrated to work in the case of baseline covariates of

interest and outcome at a given time point, and we included time-varying covariates and longitudinal outcomes. Additionally, our misspecified models omitted one of the variables in the final estimating equations, a common approach to model misspecification in prior simulation studies (Tsiatis, Davidian and Cao (2011), Seaman and Copas (2009), Schnitzer, Lok and Bosch (2016)). In our case we omitted one of the variables that defines the target estimand, a case of potentially severe model misspecification. Notably, the BR* estimator requires that the X , which define the estimand of interest, are included in the imputation models. This adds a requirement that the imputation models be tailored to the final estimand, a requirement that our imputation-based approach is designed to avoid.

This AIPW-based doubly-robust imputation framework provides a good opportunity for sensitivity analysis in randomized trials and other settings. For example, in our application to a trial of donepezil for Alzheimer's disease, it was clear that differential dropout might introduce significant bias in the comparisons of interest between study arms (Figure 1). Because the dropout rates depended on disease severity and differed significantly between the donepezil and placebo arms, it is unclear whether a standard regression model using a GEE approach would produce consistent estimates. Our application of doubly-robust imputation to the donepezil data illustrates a case where the direction of the estimated treatment effect might be reversed by using a doubly-robust estimator, indicating significant bias in the standard MMRM estimates. Indeed, our simulation studies demonstrated how a mixed-effects model with a simplified covariance matrix, such as is not infrequently used in clinical trials, can produce badly biased results. In such a situation, doubly-robust AIPW-imputation based estimators, such as AIPW-I and AIPW-S, might be helpful as a sensitivity analysis for the primary analysis. The same data might then be useful for additional analyses which is an advantage of the imputation-based approach. Importantly, the theoretical basis for consistency is well established for these DR imputation estimators in the setting of informative longitudinal dropout, which is not the case, to our knowledge, for the imputation by chained equations approach in common use. Future work will compare these doubly-robust estimators to the more usual multiple imputation approaches used as sensitivity analyses in the clinical trials setting.

Acknowledgments. The authors gratefully acknowledge Dr. Howard Feldman, PI of the Alzheimer's Disease Cooperative Study, for his mentorship and support of the first author during his Ph.D. studies at University of California San Diego.

The authors also thank the Editor, Jeffrey Morris, the corresponding Associate Editor, and two referees for insightful comments which led to important improvements over the earlier drafts.

Funding. The authors gratefully acknowledge support from grant number R01AG061146 from the US DHHS NIH National Institute on Aging, and from grant number P30 CA023100 from the NIH National Cancer Institute.

Data collection and sharing for this project were obtained from the Alzheimer's Disease Cooperative Study (ADCS), funded by the National Institutes of Health Grant U19 AG010483.

SUPPLEMENTARY MATERIAL

Supplement to "An efficient doubly-robust imputation framework for longitudinal dropout, with an application to an Alzheimer's clinical trial" (DOI: [10.1214/23-AOAS1728SUPP](https://doi.org/10.1214/23-AOAS1728SUPP); .pdf). This Supplementary Material includes four sections. Section 1 is a formal proof of the double robustness of AIPW-I estimator; Section 2 is a formal and detailed derivation of AIPW-S and demonstration of its double robustness; Section 3 is a table for simulation methods specification; Section 4 is a table for simulation results in extreme dropout construct.

REFERENCES

- BANG, H. and ROBINS, J. M. (2005). Doubly robust estimation in missing data and causal inference models. *Biometrics* **61** 962–972. MR2216189 <https://doi.org/10.1111/j.1541-0420.2005.00377.x>
- COMMITTEE FOR MEDICINAL PRODUCTS FOR HUMAN USE (2010). Guideline on Missing Data in Confirmatory Clinical Trials. London: European Medicines Agency.
- GNEITING, T. and RAFTERY, A. E. (2007). Strictly proper scoring rules, prediction, and estimation. *J. Amer. Statist. Assoc.* **102** 359–378. MR2345548 <https://doi.org/10.1198/016214506000001437>
- HSU, C.-H., HE, Y., LI, Y., LONG, Q. and FRIESE, R. (2016). Doubly robust multiple imputation using kernel-based techniques. *Biom. J.* **58** 588–606. MR3500562 <https://doi.org/10.1002/bimj.201400256>
- INTERNATIONAL COUNCIL FOR HARMONISATION OF TECHNICAL REQUIREMENTS FOR PHARMACEUTICALS FOR HUMAN USE (2017). E9(R1) Statistical Principles for Clinical Trials: Addendum: Estimands and Sensitivity Analysis in Clinical Trials. Massachusetts Medical Society.
- LIANG, K. Y. and ZEGER, S. L. (1986). Longitudinal data analysis using generalized linear models. *Biometrika* **73** 13–22. MR0836430 <https://doi.org/10.1093/biomet/73.1.13>
- LONG, Q., HSU, C.-H. and LI, Y. (2012). Doubly robust nonparametric multiple imputation for ignorable missing data. *Statist. Sinica* **22** 149–172. MR2933171 <https://doi.org/10.5705/ss.2010.069>
- PAIK, M. C. (1997). The generalized estimating equation approach when data are not missing completely at random. *J. Amer. Statist. Assoc.* **92** 1320–1329.
- PETERSEN, R. C., THOMAS, R. G., GRUNDMAN, M., BENNETT, D., DOODY, R., FERRIS, S., GALASKO, D., JIN, S., KAYE, J. et al. (2005). Vitamin E and donepezil for the treatment of mild cognitive impairment. *N. Engl. J. Med.* **352** 2379–2388.
- QIU, Y. and MESSER, K. (2023). Supplement to “An efficient doubly-robust imputation framework for longitudinal dropout, with an application to an Alzheimer’s clinical trial.” <https://doi.org/10.1214/23-AOAS1728SUPP>
- ROBINS, J. M., ROTNITZKY, A. and ZHAO, L. P. (1995). Analysis of semiparametric regression models for repeated outcomes in the presence of missing data. *J. Amer. Statist. Assoc.* **90** 106–121. MR1325118
- ROTNITZKY, A., LEI, Q., SUED, M. and ROBINS, J. M. (2012). Improved double-robust estimation in missing data and causal inference models. *Biometrika* **99** 439–456. MR2931264 <https://doi.org/10.1093/biomet/ass013>
- SAS INSTITUTE INC. (2012). Online Documentation, SAS/STAT Version 9.3: Shared Concepts: LSMEANS Statement.
- SCHARFSTEIN, D. O., ROTNITZKY, A. and ROBINS, J. M. (1999). Adjusting for nonignorable drop-out using semiparametric nonresponse models. *J. Amer. Statist. Assoc.* **94** 1096–1146. MR1731478 <https://doi.org/10.2307/2669923>
- SCHNITZER, M. E., LOK, J. J. and BOSCH, R. J. (2016). Double robust and efficient estimation of a prognostic model for events in the presence of dependent censoring. *Biostatistics* **17** 165–177. MR3449858 <https://doi.org/10.1093/biostatistics/kxv028>
- SEAMAN, S. and COPAS, A. (2009). Doubly robust generalized estimating equations for longitudinal data. *Stat. Med.* **28** 937–955. MR2518358 <https://doi.org/10.1002/sim.3520>
- SEAMAN, S. R. and VANSTEELANDT, S. (2018). Introduction to double robust methods for incomplete data. *Statist. Sci.* **33** 184–197. MR3797709 <https://doi.org/10.1214/18-STS647>
- TSIATIS, A. A. (2006). *Semiparametric Theory and Missing Data*. Springer Series in Statistics. Springer, New York. MR2233926
- TSIATIS, A. A., DAVIDIAN, M. and CAO, W. (2011). Improved doubly robust estimation when data are monotonely coarsened, with application to longitudinal studies with dropout. *Biometrics* **67** 536–545. MR2829022 <https://doi.org/10.1111/j.1541-0420.2010.01476.x>
- VAN DER LAAN, M. J. and RUBIN, D. (2006). Targeted maximum likelihood learning. *Int. J. Biostat.* **2** Art. 11. MR2306500 <https://doi.org/10.2202/1557-4679.1043>
- XU, C., LI, Z., XUE, Y., ZHANG, L. and WANG, M. (2019). An R package for model fitting, model selection and the simulation for longitudinal data with dropout missingness. *Comm. Statist. Simulation Comput.* **48** 2812–2829. MR4001237 <https://doi.org/10.1080/03610918.2018.1468457>