Robustly leveraging post-randomization information to improve precision in randomized trials

Bingkai Wang\textsuperscript{1} and Yu Du\textsuperscript{2}

\textsuperscript{1}The Statistics and Data Science Department of the Wharton School, University of Pennsylvania, PA, USA
\textsuperscript{2}Statistics, Data and Analytics, Eli Lilly and Company, IN, USA

Abstract

In randomized trials, repeated measures of the outcome are routinely collected. The mixed model for repeated measures (MMRM) leverages the information from these repeated outcome measures, and is often used for the primary analysis to estimate the average treatment effect at the primary endpoint. MMRM, however, can suffer from bias and precision loss when it models intermediate outcomes incorrectly, and hence fails to use the post-randomization information harmlessly. This paper proposes an extension of the commonly used MMRM, called IMMRM, that improves the robustness and optimizes the precision gain from covariate adjustment, stratified randomization, and adjustment for intermediate outcome measures. Under regularity conditions and missing completely at random, we prove that the IMMRM estimator for the average treatment effect is robust to arbitrary model misspecification and is asymptotically equal or more precise than the analysis of covariance (ANCOVA) estimator and the MMRM estimator. Under missing at random, IMMRM is less likely to be misspecified than MMRM, and we demonstrate via simulation studies that IMMRM continues to have less bias and smaller variance. Our results are further supported by a re-analysis of a randomized trial for the treatment of diabetes.

Keywords: ANCOVA, heterogeneity, heteroscedasticity, MMRM, repeated outcome measures, robustness
1 Introduction

“Intermediate outcomes” in the context of randomized trials refer to the outcomes measured after treatment assignment but before the time point of interest. For example, for a trial studying the effect of a 12-month dietary plan on weight, the intermediate outcomes can be the weights at 3, 6, and 9 months after randomization. Intermediate outcomes are routinely collected and have been used in various ways, including trial monitoring (Shih and Quan, 1999), decision making in the interim analysis (Kunz et al., 2015), mediation analysis (Landau et al., 2018), etc. We focus on a common but less studied purpose of intermediate outcomes, which is to robustly improve the precision of statistical inference on the average treatment effect at the primary endpoint.

The status quo for improving the statistical precision of longitudinal trials is adjusting for covariates and intermediate outcomes, for which the mixed model for repeated measures (MMRM, Mallinckrodt et al., 2008) is commonly used. MMRM involves a linear mixed model for the outcome measure at each visit on fixed effects (including the intercept, treatment indicators, and covariates, where the coefficients for covariates are often modeled as constant across visits) and a normal-distributed random intercept (which characterizes the correlation across outcome measures). If MMRM correctly specifies both fixed and random effects, it will yield consistent estimates for the average treatment effect at the primary endpoint and can improve precision by adjusting for intermediate outcomes (Marschner and Becker, 2001; Galbraith and Marschner, 2003; Stallard, 2010; Hampson and Jennison, 2013; Zhou et al., 2018). However, no analytical results are given when MMRM is arbitrarily misspecified. For example, it has been concerned that the outcomes of a randomized trial can be non-linear on covariates or non-normally distributed (Kraemer, 2015); in this case, MMRM may not capture the true relationship among outcomes, treatment, and covariates, thereby leaving open questions related to validity and precision of a misspecified MMRM.

In this article, we study whether a misspecified MMRM remains valid and precise. Under
mild regularity conditions, we prove that the MMRM estimator for the average treatment effect at the primary endpoint is “robust”, i.e., it is consistent and asymptotically normal under arbitrary model misspecification, if there are no missing outcomes or the missingness is completely at random. This result also holds under stratified randomization (Zelen, 1974) or given multiple treatments. Under missing at random, however, a misspecified MMRM could lead to bias. In terms of precision, we contribute a surprising result that adjustment for intermediate outcomes by a misspecified MMRM could result in precision loss, even when visit-by-covariates interaction terms are incorporated. This finding is supported by both theoretical derivations and simulations, indicating that the current MMRM may adjust for post-randomization information inappropriately and inadequately.

To address the above issues of MMRM, we propose “Improved MMRM” (IMMRM), an extension of MMRM that could augment robustness and precision by modeling treatment heterogeneity and heteroscedasticity. We show that IMMRM not only shares the same robustness property discussed above with MMRM, but is also less likely to be misspecified since MMRM is a special case of IMMRM, thereby increasing its reliability under the missing at random assumption. In terms of precision, IMMRM is the first model, to the best of our knowledge, that combines the precision gain from covariate adjustment, stratified randomization, and adjustment for intermediate outcomes. We prove that, under missing completely at random, (1) IMMRM is asymptotically equally or more precise than MMRM, (2) adjustment for intermediate outcomes by IMMRM will retain or reduce, but not increase the asymptotic variance, and (3) IMMRM could automatically accommodate the precision gain from stratified randomization. In addition, we formally characterize when stratified randomization and intermediate outcomes improve precision beyond covariate adjustment. Computationally, we demonstrate via simulation studies that this precision gain also exists for data with a small sample size under the missing at random assumption.

Our results build on existing work on causal inference under the super-population framework. For the validity of misspecified MMRM and IMMRM, our proof is based on the
semiparametric theory (Tsiatis, 2007) and its recent extension to stratified randomization (Bugni et al., 2019). In the research area of mixed models, although misspecified random effects have been extensively studied (McCulloch and Neuhaus, 2011), the existing results mainly relied on correctly-specified fixed effects and focused on random effects estimation, which is not the topic of this article. For robustly improving precision by adjusting for intermediate outcomes, Qian et al. (2019) used the general approach proposed by Lu and Tsiatis (2011); van der Laan and Gruber (2012) and derived the non-parametric formulas for the precision gain from adjusting for baseline variables and the short-term outcomes. Later, Van Lancker et al. (2020) proposed a robust estimator that achieves such precision gain for the interim analysis. More recently, Schuler (2022) showed the analysis of covariance, which ignores intermediate outcomes, outperforms MMRM (without visit-by-covariate interaction terms) under two-arm, equal, simple randomization with no missing data. However, the above results fail to handle the complexity of non-monotone censoring at multiple visits or account for stratified randomization. Finally, our results are extensions of established work on variance reduction by covariate adjustment (Yang and Tsiatis, 2001; Rubin and van der Laan, 2008; Jiang et al., 2018, among others) and stratified randomization (Ye et al., 2021; Wang et al., 2021, among others).

In the next section, we introduce the motivating example. In Section 3, we present our setup, notations, and assumptions. In Section 4, we describe the ANCOVA estimator, MMRM estimator (with or without visit-by-covariates interactions), and our proposed IMMRM estimator. Our main result is presented in Section 5, which consists of asymptotic theory, explanations of precision gain from intermediate outcomes, and results for two-arm equal randomization. Simulation study and data application are given in Sections 6 and 7, respectively. Finally, we provide practical recommendations and discuss future directions in Section 8.
2 Motivating example the IMAGINE-2 study

The trial of “A Study in Patients With Type 2 Diabetes Mellitus (IMAGINE 2)” is a 52-week, two-arm, phase 3 randomized clinical study completed in 2014 (Davies et al., 2016). The goal of this trial was to evaluate the effect of basal insulin for the treatment of type 2 diabetes in an insulin-naïve population.

Participants were randomly assigned to receive insulin peglispro (treatment, 1003 patients) or insulin glargine (control, 535 patients), aiming to achieve a 2:1 randomization ratio. Randomization was stratified by baseline HbA1c (Hemoglobin A1c, < 8.5% or ≥ 8.5%), low-density lipoprotein cholesterol (< 100mg/dL or ≥ 100mg/dL) and baseline sulphonylurea or meglitinide use (yes or not). HbA1c is a continuous measure of average blood glucose values in the prior three months. The primary outcome was the change in HbA1c at week 52 from baseline (15% missing outcomes), while intermediate outcomes were measured at week 4 (3% missing outcomes), week 12 (6% missing outcomes), week 26 (10% missing outcomes) and week 39 (13% missing outcomes). We focus on estimating the average treatment effect of the primary outcome and adjusting for intermediate outcomes, baseline HbA1c value, and randomization strata.

3 Definition and assumptions

3.1 Data generating distributions

Consider a trial where the outcome is continuous and repeatedly measured at $K$ visits, where $K$ is a positive integer. For each participant $i = 1, \ldots, n$, the outcome vector at $K$ visits is $Y_i = (Y_{i1}, \ldots, Y_{iK}) \in \mathbb{R}^K$ and the non-missing status at $K$ visits is $M_i = (M_{i1}, \ldots, M_{iK})$, where $M_{it} = 1$ if $Y_{it}$ is observed at visit $t$, and 0 otherwise. Let $A_i$ be a categorical variable taking values in $\{0, 1, \ldots, J : J \geq 1\}$, with $A_i = j$ representing that participant $i$ is assigned to the $j$-th treatment group. By convention, we use $A_i = 0$ to denote being assigned to the
control group. Let \( X_i \) be a vector of baseline variables with length \( p \). Throughout, we refer to \( Y_K \) as the final outcome, and \( (Y_1, \ldots, Y_{K-1}) \) as the intermediate outcomes (if \( K > 1 \)), for conciseness.

We use the Neyman-Rubin potential outcomes framework (Neyman et al., 1990), which assumes \( Y_i = \sum_{j=0}^{J} I\{A_i = j\} Y_i(j) \), where \( I \) is the indicator function and \( Y_i(j) = (Y_{i1}(j), \ldots, Y_{iK}(j)) \) is the potential outcome vector for treatment group \( j = 0, \ldots, J \). Analogously to the consistency assumption above, we assume \( M_i = \sum_{j=0}^{J} I\{A_i = j\} M_i(j) \), where \( M_i(j) \) is the indicator vector of whether \( Y_i(j) \) would be observed at each of the \( K \) visits, if participant \( i \) was assigned to treatment group \( j \) for \( j = 0, \ldots, J \).

For each participant \( i \), we define the complete data vector as \( W_i = (Y_i(0), \ldots, Y_i(J), M_i(0), \ldots, M_i(J), X_i) \) and the observed data vector as \( O_i = (Y_i^o, M_i, A_i, X_i) \), where \( Y_i^o \) is the vector of observed outcomes, whose dimension may vary across participants. For example, if participant \( i \) only shows up in visits 1 and \( K \), then \( Y_i^o = (Y_{i1}, Y_{iK}) \). For the special case that a participant misses all post-randomization visits but still has baseline information recorded, the observed data vector is \( O_i = (M_i, A_i, X_i) \) with \( M_i \) being a zero vector. All estimators defined in Section 4 are functions of the observed data \((O_1, \ldots, O_n)\). We make the following assumptions on \((W_1, \ldots, W_n)\):

**Assumption 1.**

(a) \( W_i, i = 1, \ldots, n \) are independent, identically distributed samples from the joint distribution \( P \) on \( W = (Y(0), \ldots, Y(J), M(0), \ldots, M(J), X) \).

(b) (Missing completely at random, MCAR) \( M(j) \) is independent of \( (Y(j), X) \) for \( j = 0, \ldots, J \), and \( M(0), \ldots, M(J) \) are identically distributed.

(c) (Positivity) \( P(M_1(j) = 1, \ldots, M_K(j) = 1) > 0 \).

Assumption 1 (a) is the standard assumption for causal inference under the superpopulation framework. Assumption 1 (b) implies that the missing pattern is independent of
outcomes and covariates. This assumption, while restrictive, is needed for achieving valid inference under arbitrary model misspecification (See Section 5). An alternative and more flexible assumption is missing at random (MAR), i.e., $M_t(j)$ is independent of $Y_t(j)$ given \{X, M_1(j), Y_1(j), \ldots, M_{t-1}(j), Y_{t-1}(j)\}. Under MAR, however, all models we examine in Section 4 could be biased if they are misspecified, and we empirically test their performance via simulation studies (Section 6). Finally, Assumption 1 (c) requires the existence of “completers” in each arm, which is a convenient assumption for achieving parameter identifiability in the considered models.

In addition to Assumption 1, we also assume regularity conditions for the estimators defined in Section 4. As we show in the Supplementary Material, all estimators we consider (including our proposed estimators) are M-estimators (Section 5 of van der Vaart, 1998), which is defined as a zero of prespecified estimating functions. These conditions are similar to the classical non-parametric conditions given in Section 5.3 of van der Vaart (1998) for proving consistency and asymptotic normality for M-estimators. We provide these regularity conditions in the Supplementary Material.

The parameters of interest are the average treatment effects of the final outcome comparing each treatment group to the control group, i.e.,

$$\Delta^* = (E[Y_K(1)] - E[Y_K(0)], \ldots, E[Y_K(J)] - E[Y_K(0)]),$$

where $E$ is the expectation with respect to the distribution $P$. Our results in Section 5 also apply to estimating any linear transformation of $\Delta^*$, e.g., the average treatment effect comparing any two treatment groups.

### 3.2 Simple and stratified randomization

We consider two types of treatment assignment procedures: simple randomization and stratified randomization. For $j = 0, \ldots, J$, let $\pi_j$ be the target proportion of participants assigned
to the treatment group $j$. We assume that $\sum_{j=0}^{J} \pi_j = 1$ and $\pi_j > 0$ for all treatment groups. For example, equal randomization refers to the setting where $\pi_0 = \cdots = \pi_J$.

Simple randomization allocates treatment by independent draws from a categorical distribution on $A$, with $P(A = j) = \pi_j$ for $j = 0, \ldots, J$. Then $(A_1, \ldots, A_n)$ are independent, identically distributed samples from this categorical distribution, and also independent of $(W_1, \ldots, W_n)$.

Under stratified randomization, treatment assignment depends on a set of categorical baseline variables, called stratification variables. We use a categorical random variable $S$ with support $S = \{1, \ldots, R\}$ to denote the joint levels created by all stratification variables. For example, if randomization is stratified by sex (female or male) and weight (normal, overweight, or obesity), then $S$ can take $R = 6$ possible values. Each element in $S$ is referred to as a “randomization stratum”. Within each randomization stratum, permuted blocks are used for sequential treatment allocation. Within each permuted block, fraction $\pi_j$ of participants are assigned to treatment group $j$. After a block is exhausted, a new block is used. For each participant $i$, we assume that $S_i$ is encoded as $R - 1$ dummy variables (dropping one level to avoid collinearity) in the baseline vector $X_i$.

Different from simple randomization, stratified randomization is able to achieve balance within each randomization stratum, i.e., exact fraction $\pi_j$ of participants are assigned to group $j$. Due to this advantage, stratified randomization is used by 70% of randomized clinical trials, according to a survey by Lin et al. (2015). In the meantime, stratified randomization also brings statistical challenge: the treatment assignments $(A_1, \ldots, A_n)$ are not independent of each other such that the classical central limit theory cannot be directly applied. To overcome this challenge, recent work (Ye et al., 2021; Wang et al., 2021 among others) has established the non-parametric asymptotic theory for common estimators in non-longitudinal randomized trials. However, the analysis of robustness and efficiency for longitudinal trials, to the best of our knowledge, remains incomplete under stratified randomization.
We finish this section by introducing a few additional definitions. For any two symmetric matrices $V_1$ and $V_2$ with the same dimension, we denote $V_1 \succeq V_2$ if $V_1 - V_2$ is positive semi-definite. For any estimator $\hat{\Delta}$ of $\Delta^*$, we call $V$ the asymptotic covariance matrix of $\hat{\Delta}$ if $\sqrt{n}(\hat{\Delta}_1 - \Delta^*)$ weakly converges to a multivariate normal distribution with mean 0 and covariance matrix $V$. If two estimators $\hat{\Delta}_1$ and $\hat{\Delta}_2$ of $\Delta^*$ have asymptotic covariance matrix $V_1$ and $V_2$ respectively, we call $\hat{\Delta}_2$ is (asymptotically) equally or more precise than $\hat{\Delta}_1$ if $V_1 \succeq V_2$. Such an expression is commonly used for scalar estimators, and we extend it to vector estimators.

4 Estimators

For estimating the average treatment effects $\Delta^*$, we introduce three commonly used models in Sections 4.1-4.2 and then propose a new working model in Section 4.3. We focus on comparing these four models in the rest of this article.

4.1 The Analysis of Covariance (ANCOVA) estimator

The ANCOVA estimator, $\hat{\Delta}^{(\text{ANCOVA})}$, is defined as the maximum likelihood estimator (MLE) for parameters $(\beta_{A1}, \ldots, \beta_{AJ}, \beta_X)$ in the working model

$$Y_K = \beta_0 + \sum_{j=1}^J \beta_{Aj} I\{A = j\} + \beta_X^T X + \varepsilon, \quad (1)$$

where $(\beta_0, \beta_{A1}, \ldots, \beta_{AJ}, \beta_X)$ are parameters, and $\varepsilon$ is independent of $X$ and follows a normal distribution with mean 0 and unknown variance $\sigma^2$. Participants with missing outcomes will be dropped from the model fit. When outcomes are repeatedly measured, the ANCOVA estimator wastes the information from intermediate outcomes. Although ignoring such information does not affect the robustness of ANCOVA under MCAR (Yang and Tsiatis, 2001), intermediate outcomes can be used to improve precision, as we show below.
4.2 The mixed-effects model for repeatedly measured outcomes (MMRM)

There are two MMRM working models that are commonly seen in the analysis of randomized trials: one adjusts for covariates by a constant coefficient vector across visits, while the other includes visit-by-covariates interaction terms. For conciseness, we refer to the former model as “MMRM-I” and the latter as “MMRM-II”.

The MMRM-I working model is defined as, for each $t = 1, \ldots, K$,

$$Y_t = \beta_{0t} + \sum_{j=1}^{J} \beta_{Aj} I\{A = j\} + \beta_{X}^T X + \varepsilon_t,$$

where $\{(\beta_{0t}, \beta_{A1t}, \ldots, \beta_{AJt})\}_{t=1}^{K}$ and $\beta_{X}$ are parameters, and $\varepsilon = (\varepsilon_1, \ldots, \varepsilon_K)$ is the independent residual error following a multivariate normal distribution with mean $0$ and unknown covariance $\Sigma$. The covariance matrix $\Sigma$ is assumed to be unstructured; that is, no other assumption is made on $\Sigma$ except that it is positive definite. The MMRM-I estimator $\hat{\Delta}^{\text{MMRM}}$ for $\Delta^*$ is defined as the MLE for parameters $(\beta_{A1K}, \ldots, \beta_{AJK})$. Of note, Equation (2) is the marginal model, i.e., the random effects are marginalized and implicitly represented in the covariance matrix $\Sigma$.

The MMRM-II working model is, for each $t = 1, \ldots, K$,

$$Y_t = \beta_{0t} + \sum_{j=1}^{J} \beta_{Aj} I\{A = j\} + \beta_{X_t}^T X + \varepsilon_t,$$

which differs from the MMRM-I working model only on the regression coefficients of $X$, where $\beta_{X_t}$ is substituted for $\beta_{X}$. We define the MMRM-II estimator, $\hat{\Delta}^{\text{MMRM-II}}$, as the MLE for $(\beta_{A1K}, \ldots, \beta_{AJK})$ in model (3). Different from MMRM-I, MMRM-II can capture the time-varying correlation of covariates and outcomes and has been recommended for the primary analysis (Mallinckrodt et al., 2020).

In MMRM-I (2) and MMRM-II (3), the correlation of covariates $X$ and the outcome vector $Y$ is constant among treatment groups, which we call the homogeneity assumption. In
addition, the covariance matrix of $Y$ is also assumed to be the same across treatment groups, which we refer to as the homoscedasticity assumption. Due to these assumptions, although MMRM-I and MMRM-II utilize information from intermediate outcomes, we find that they may be less precise than ANCOVA if their working models are misspecified (Section 5.2).

### 4.3 Improved MMRM: modeling heterogeneity and heteroscedasticity among groups and visits

We propose a working model, called “IMMRM”, that handles both heterogeneity and heteroscedasticity as follows: for each $t = 1, \ldots, K$,

$$Y_t = \beta_0 + \sum_{j=1}^J \beta_{Ajt} I\{A = j\} + \beta_t^T X + \sum_{j=1}^J \beta_{AXjt}^T I\{A = j\} X + \sum_{j=0}^J \varepsilon_j I\{A = j\},$$

where $\varepsilon_j = (\varepsilon_{j1}, \ldots, \varepsilon_{jK})$ has a multivariate normal distribution with mean 0 and covariance $\Sigma_j$ for each $j = 0, \ldots, J$, and $(\varepsilon_0, \ldots, \varepsilon_J)$ are independent of $X$ and each other. The fixed effects in model (4) are $(\beta_0, \beta_{Ajt}, \beta_{Xjt}, \beta_{AXjt})$ for $j = 1, \ldots, J$. Each $\Sigma_j, j = 0, \ldots, J$ is assumed to be positive definite and unstructured. The IMMRM estimator for $\Delta^*$ is defined as

$$\hat{\Delta}^{(IMMRM)} = (\hat{\beta}_{A1K} + \hat{\beta}_{AX1K}^T \overline{X}, \ldots, \hat{\beta}_{AJK} + \hat{\beta}_{AXJK}^T \overline{X}),$$

where $\overline{X} = n^{-1} \sum_{i=1}^n X_i$ and $(\hat{\beta}_{AJK}, \hat{\beta}_{AXJK})$ are MLEs for parameters $(\beta_{AJK}, \beta_{AXJK})$.

Compared with MMRM-I and MMRM-II, the IMMRM working model has two improvements. First, the inclusion of treatment-covariates-visits three-way interaction terms allows the relationship between the outcomes and baseline variables to vary across treatment groups and visits. Such interaction terms models heterogeneity, which has been shown by Tsiatis (2007); Ye et al. (2021) as an effective method to improve precision for scalar outcomes. We extend this idea to longitudinal repeated measures data. Second, the covariance matrix of $Y$ is modeled separately for each treatment group, which accounts for heteroscedasticity. Gosho and Maruo (2018) first proposed the idea of modeling heteroscedasticity in MMRM;
however, they only provide empirical results to show its benefits. We show, in Section 5.2, that modeling heteroscedasticity is necessary for achieving asymptotic precision gain when repeated measure outcomes are jointly modeled.

Technically, IMMRM remains a linear mixed model for repeated measures and hence fits into the broad MMRM framework defined by Mallinckrodt et al. (2008). However, IMMRM extends the commonly used MMRM working models, i.e., MMRM-I and MMRM-II, by handling heterogeneity and heteroscedasticity, two techniques that have not been used together in practice or theoretically analyzed for longitudinal randomized trials, thereby representing a methodology improvement with a practical impact.

5 Main results

5.1 Asymptotic theory

Theorem 1. Assume Assumption 1 and regularity conditions. Consider $\hat{\Delta}^{(est)}$ for est $\in \{\text{ANCOVA, MMRM-I, MMRM-II, IMMRM}\}$.

For each of the four estimators, under simple or stratified randomization, we have consistency, i.e., $\hat{\Delta}^{(est)} \rightarrow \Delta^*$ in probability, and asymptotic normality, i.e., $\sqrt{n}(\hat{\Delta}^{(est)} - \Delta^*)$ weakly converges to a mean-zero multivariate normal distribution, under arbitrary misspecification of its working model.

Denote $\tilde{V}^{(est)}$ and $V^{(est)}$ as the asymptotic covariance matrices of $\hat{\Delta}^{(est)}$ under simple and stratified randomization, respectively. We have, for each est $\in \{\text{ANCOVA, MMRM-I, MMRM-II}\}$,

$$\tilde{V}^{(est)} \succeq V^{(est)} \succeq \tilde{V}^{(IMMRM)} = V^{(IMMRM)}. \quad (5)$$

In addition, we provide the conditions for $V^{(est)} = V^{(IMMRM)}$ in the Supplementary Material.

Theorem 1 has the following implications. First, under simple or stratified randomization, each of the ANCOVA, MMRM-I, MMRM-II, and IMMRM estimators is robust under
MCAR. Second, the IMMRM estimator has the highest precision among the four estimators. Such precision gain can be translated into sample size reduction. Third, unlike the other three estimators, the IMMRM estimator has the same asymptotic covariance matrix under simple or stratified randomization. Therefore, the confidence interval for the IMMRM estimator under stratified randomization can be constructed as if simple randomization were used, without being statistically conservative.

For performing hypothesis testing and constructing confidence intervals, we provide consistent estimators for the asymptotic covariance matrices $\tilde{V}^{(est)}$ and $V^{(est)}$ in the Supplementary Material. The sandwich variance estimator (Tsiatis, 2007) is used to estimate $\tilde{V}^{(est)}$; and the expression of $\tilde{V}^{(est)} − V^{(est)}$ is derived in the Supplementary Material and approximated by substituting $\hat{E}$, the expectation with respect to the empirical distribution, for $E$.

Our proof for Theorem 1 is provided in the Supplementary Material. In the proof, we derive the influence function for each of the four estimators and prove the asymptotic results by extending Theorem 1 of Wang et al. (2021) and Lemma B.2 of Bugni et al. (2019) to longitudinal data. The major innovation and challenge of the proof are to derive the partial order (5) given the non-monotone missingness. We overcome this challenge by developing a series of inequalities related to functions of $M$ and positive definite matrices, which are presented in Lemma 1 of the Supplementary Material.

When the missing data mechanism depends on the treatment, baseline information, or historical outcomes, all estimators considered in Theorem 1 could be biased if their working models are misspecified, a common property shared by many outcome regression models (Wang et al., 2019). However, since MMRM-I and MMRM-II are special cases of IMMRM, IMMRM has additional robustness to model misspecification: it can still be correct when MMRM-I and MMRM-II are wrong, but not vice versa. Furthermore, the sensitivity analysis on missing data assumptions for MMRM (O’Kelly and Ratitch, 2014; Mallinckrodt et al., 2020) can also be applied to IMMRM, allowing for its incorporation into the current practice.
In the ensuing real-data-based simulation study, we show that the IMMRM estimator is less biased and more precise than the others under MAR.

5.2 How adjustment for intermediate outcomes improves precision

Consider the IMMRM working model (4) with a modification that intermediate outcomes are excluded, i.e.,

\[ Y_K = \beta_{0K} + \sum_{j=1}^{J} \beta_{jK} I\{A = j\} + \beta_{XK}^\top X + \sum_{j=1}^{J} \beta_{jA} K I\{A = j\} X + \sum_{j=1}^{J} \varepsilon_{jK} I\{A = j\}. \tag{6} \]

Then the IMMRM estimator for \( \Delta^* \) given the above working model (6) reduces to the “ANHECOVA” estimator proposed by Ye et al. (2021). As a special case of the IMMRM model (4), the working model (6) differs from IMMRM only on whether intermediate outcomes are adjusted. By comparing the asymptotic covariance matrices of the IMMRM estimator with the ANHECOVA estimator, we examine the contribution of intermediate outcomes in improving precision beyond what comes from covariate adjustment and stratified randomization.

If \( K = 1 \), a case with no intermediate outcomes, the IMMRM estimator and ANHECOVA estimators are equivalent. For \( K > 2 \), the following Corollary shows that adjusting for intermediate outcomes by IMMRM will retain or increase but not lose precision.

**Corollary 1.** Assume \( K > 1 \), Assumption 1 and regularity conditions in the Supplementary Material. Let \( V^{(ANHECOVA)} \) be the asymptotic covariance matrix of the ANHECOVA estimator based on the working model (6). Then \( V^{(ANHECOVA)} \succeq V^{(IMMRM)} \).

Furthermore, \( V^{(ANHECOVA)} = V^{(IMMRM)} \) if and only if, for each \( t = 1, \ldots, K - 1 \) and \( j = 0, \ldots, J \),

\[ P(M_t(j) = 1, M_K(j) = 0) \text{ Cov}\{Y_t(j), Y_K(j) - \beta_{Kj}^\top X\} = 0, \]
where $b_{Kj} = \text{Var}(X)^{-1}\text{Cov}\{X, Y_{K}(j)\}$, $\text{Cov}\{U_1, U_2\} = E[U_1U_1^\top] - E[U_1]E[U_1]^\top$ is the covariance between any random vectors $U_1, U_2$ with finite second moments, and $\text{Var}(X) = \text{Cov}(X, X)$ is the covariance matrix of $X$.

Corollary 1 specifies the conditions for when adjusting for intermediate outcomes brings precision gain. For an intermediate visit $t$, $P(M_t(j) = 1, M_K(j) = 0) > 0$ implies that a participant has a positive probability to both appear in visit $t$ and miss the last visit $K$; and $\text{Cov}\{Y_t(j), Y_{K}(j) - b_{Kj}^\top X\} \neq 0$ means that, for the treatment group $j$, $Y_t(j)$ is correlated with the residual of $Y_K(j)$ after regressing on baseline variables. If an intermediate outcome satisfies the above two conditions for some treatment group $j$, then it will lead to precision gain. On the contrary, adjusting for an intermediate outcome makes no change on the asymptotic covariance matrix if an intermediate outcome $Y_t$ is missing whenever $Y_K$ is missing, or if it is not prognostic to the final outcome after controlling for $X$ in any treatment group.

Leveraging intermediate outcomes can bring precision gain only when there are missing final outcomes and the intermediate outcomes are prognostic to the final outcome beyond what is explained by covariates. This result also applies to MMRM-I and MMRM-II. When either of the two condition fails, ANHECOVA will lead to the same precision as IMMRM and be equally or more precise than MMRM-I and MMRM-II. This finding generalizes the results of Qian et al. (2019), which considers a special case of our setup with $J = K = 2$, simple randomization, and monotone censoring.

Unlike the IMMRM estimator, adjusting for intermediate outcomes by MMRM-I or MMRM-II may cause efficiency loss. MMRM-I uses a constant vector to model the effect of covariates on outcomes, which is too stringent to capture its variation across visits and treatment groups. Thus, it can be less precise than ANCOVA, which only models the final outcome. This result is demonstrated in both simulation study and data application. For the MMRM-II estimator, although simulation studies have shown that it is compara-
ble to or better than ANCOVA (Mallinckrodt et al., 2020; Schuler, 2022), our theoretical examination indicates that it is not universally correct: MMRM-II can be less precise than ANCOVA when there is heteroscedasticity. To make this argument concrete, we construct an analytical counterexample in the supplementary material showing that the MMRM-II estimator has a 5% larger variance than the ANCOVA estimator. Of note, an exception is the two-arm equal randomization, where MMRM-II is asymptotically equally or more precise than ANCOVA, as discussed below.

5.3 Special case: two-arm equal randomization

In a general setting, e.g., multi-arm trials or unequal randomization, the efficiency comparison among ANCOVA, MMRM-I, and MMRM-II is indeterminate. Under the two-arm equal randomization, the following corollary implies that the MMRM-II estimator has equal or smaller asymptotic variance than ANCOVA and MMRM-I; in addition, the MMRM-II estimator has the same asymptotic variance under simple or stratified randomization.

**Corollary 2.** Assume $J = 1$, $\pi_1 = \pi_0$, Assumption 1, and regularity conditions in the Supplementary Material. Then $V^{(ANCOVA)} \geq V^{(MMRM-II)}$, $V^{(MMRM-I)} \geq V^{(MMRM-II)}$ and $\tilde{V}^{(MMRM-II)} = V^{(MMRM-II)}$.

6 Simulation study

6.1 Simulation settings

We conducted a simulation study assessing the performance of the ANCOVA, MMRM-I, MMRM-II and IMMRM estimators in four scenarios: small (50) versus large (200) sample size per arm, and MCAR versus MAR. Across all scenarios, the simulated data were based on the IMAGINE-2 study introduced in Section 2, reflecting a real-world data distribution.
and indicating that all working models are potentially misspecified. Under the setting of large sample size and MCAR, we demonstrate our results in Section 5. For small sample size or MAR, we stress-test the performance of all four estimators.

In the simulation, we considered five post-randomization visits \((K = 5)\), three treatment groups \((J = 2)\) and four randomization strata \((R = 4)\), trying to duplicate the setting of the IMAGINE-2 study while considering a multi-arm study. We used three baseline variables from the data: baseline HbA1c \((X_{1i})\), indicator of LDL cholesterol \(\geq 2.6\text{mmol/L} (X_{2i})\), and indicator of baseline HbA1c \(\geq 8.5\% (X_{3i})\). The simulated data were generated by the following steps.

First, we took the 844 completers from the insulin peglispro arm to serve as the super-population for the control group, where the potential outcome vector for participant \(i\) was denoted as \(Y_{i}(0)\). For the other two treatment groups, named TRT1 and TRT2, we generated the potential outcome \(Y_{i}(1)\) and \(Y_{i}(2)\) by:

\[
Y_{it}(j) = c_{t}(j) + Y_{it}(0) + \alpha_t(j)(X_{1i} - \overline{X_{1i}}) + \gamma_t(j)(X_{2i}^2 - \overline{X_{2i}^2}),
\]

where, for \(j = 1, 2\) and \(t = 1, \ldots, K\), \(c_{t}(j)\) is a constant specifying the average treatment effect comparing TRTa to the control group at time \(t\), \(\alpha_t(j)\) and \(\gamma_t(j)\) are coefficients that determine the degree of heterogeneity among treatment arms, and \(\overline{X_{1i}}\) and \(\overline{X_{2i}^2}\) are averages of \(X_{1i}\) and \(X_{2i}^2\) across \(i\) in the control group. The quadratic terms adds another layer of model misspecification. We set \(c(1) = (0, 0, 0, 0, 0)^\top\), \(c(2) = (-0.2, -0.5, -0.8, -0.9, -1)^\top\). This indicates that the true average treatment effect is \(\Delta^* = (0, -1)\). We let \(\alpha(1) = -10^{-2} \times (1, 2, 4, 5, 5)^\top\), \(\alpha(2) = -10^{-2} \times (0.5, 1, 1.5, 7, 10)^\top\), \(\gamma(1) = -10^{-2} \times (1, 3, 3, 2, 5)^\top\) and \(\gamma(2) = -10^{-2} \times (1.5, 2, 1, 10, 10)^\top\). These values were set to explore a range of mild treatment heterogeneity. The negative signs of \(\alpha(j)\) and \(\gamma(j)\) indicated that a higher baseline HbA1c is associated with a larger HbA1c change.

Next, we independently resample \(n\) participants with replacement from the empirical distribution of \(\{(Y_{i}(0), Y_{i}(1), Y_{i}(2), X_{1i}, X_{2i}, X_{3i})\}_{i=1}^{844}\). We then applied stratified random-
ization with a block size of 6 to randomly assign the resampled participants to three treatment arms with 1:1:1 randomization ratio, where randomization strata are defined by the joint levels of $X_{2i}$ and $X_{3i}$. The treatment variable was $A_i$ taking values in \{0, 1, 2\} and the realized outcome vector was $Y_i = \sum_{j=0}^{2} I\{A_i = j\}Y_i(j)$.

In the final step, we assigned missing outcomes under monotone censoring. We mimicked the missing data percentages in the IMAGINE-2 study such that 3%, 6%, 10%, 13%, and 15% are expected to be missing at visits 1-5, respectively. For MCAR, the censoring time was generated by a logistic regression (with an intercept only) to achieve the above missing data percentages. For MAR, the censoring time was determined by a logistic regression model on the treatment group, previous HbA1c value, and baseline HbA1c. The dropout rate was made higher in the control arm and TRT1 compared to TRT2, and the participants are more likely to drop out given a higher HbA1c observed from the previous visit. The complete model specifications for MAR is provided in the Supplementary Material.

Given a simulated data set, all four estimators were used to estimate $\Delta^*$. The standard error is calculated by the sandwich variance estimator (using the option “empirical” in SAS with an adjustment for the variability in the mean covariates as pointed out by Qu and Luo, 2015 in estimating the average treatment effect), which does not account for stratified randomization. The 95% confidence interval was constructed by normal approximations. The above procedure was repeated 10,000 times for each scenario. The performance metrics are the bias, empirical standard error, averaged standard error for each estimate, coverage probability, probability of rejecting the null hypothesis, and the relative mean squared error compared to IMMRM.

### 6.2 Simulation results

The simulation results are summarized in Table 1 for $n = 150$ and Table 2 for $n = 600$. Across all scenarios, IMMRM maintains a coverage probability close to 95% and consistently
outperforms the other estimators in bias and precision, reflected by its smallest RMSE. Furthermore, ANCOVA, MMRM-I, and MMRM-II fail to account for the precision gain from stratified randomization, resulting in overestimating the true standard error (comparing ASE and ESE) and thus losing power.

Under MCAR, all estimators have negligible bias despite their misspecified models, which is consistent with our asymptotic result. However, a small sample size leads to slightly larger bias and less accurate standard error estimates than the large sample setting due to increased uncertainty. The MMRM-I estimator has the largest mean squared error and variance among the four estimators, indicating its inadequate adjustment for intermediate outcomes. When the treatment has no effect (TRT1), IMMRM and MMRM-II have similar performance and are 14% more precise than ANCOVA. However, given a non-zero treatment effect (TRT2), IMMRM is 15% more precise than both ANCOVA and MMRM-II. In addition, ANCOVA, MMRM-I, and MMRM-II have an additional 4-35% variance increase due to ignoring stratified randomization, whereas IMMRM remains anti-conservative. These results further support our theorems in Section 5.

Compared to MCAR, the MAR setting increases the bias of ANCOVA and MMRM-II, while the performance of IMMRM and MMRM-I remains unchanged. In our simulation, the IMMRM estimator remains unbiased and precise under MAR, a combined result of its augmented robustness and the relatively small proportion of missing outcomes (3-15% across visits). Our choice of missing outcome proportions duplicates the IMAGINE-2 study, thereby testing the performance of IMMRM in a real-world setting (phase 3 clinical trial). If more outcomes were missing (which could occur in a different study), the IMMRM estimator, as well as the other three estimators, could have increased bias and variance.
Table 1: Simulation results comparing candidate estimators with 50 samples per arm under MCAR and MAR. For each estimator, we estimate the average treatment effect of TRT1 and TRT2, both comparing the control group. The following measures are used: bias, empirical standard error (ESE), averaged standard error (ASE), coverage probability (CP), probability of rejecting the null (PoR), relative mean squared error compared to IMMRM (RMSE). For RMSE, a number bigger than 1 indicates a larger mean squared error than IMMRM.

| Group  | Bias  | ESE  | ASE  | CP(%) | PoR(%) | RMSE |
|--------|-------|------|------|-------|--------|------|
| MCAR   |       |      |      |       |        |      |
| ANCOVA | TRT1  | 0.004| 0.229| 0.238 | 95.7   | 4.3  | 1.091|
|        | TRT2  | 0.007| 0.269| 0.275 | 95.1   | 96.0 | 1.088|
|        | MMRM-I| TRT1 | 0.006| 0.232| 0.264 | 97.2 | 2.8  | 1.105|
|        | TRT2  | 0.004| 0.301| 0.370 | 98.5   | 82.1 | 1.219|
|        | MMRM-II| TRT1| 0.005| 0.216| 0.212 | 94.6 | 5.4  | 1.028|
|        | TRT2  | 0.005| 0.268| 0.284 | 96.0   | 95.6 | 1.086|
|        | IMMRM | TRT1 | 0.002| 0.210| 0.199 | 93.1 | 6.9  | 1.000|
|        | TRT2  | 0.003| 0.247| 0.239 | 93.6   | 98.7 | 1.000|
| MAR    |       |      |      |       |        |      |
| ANCOVA | TRT1  | 0.018| 0.228| 0.239 | 96.1   | 3.9  | 1.082|
|        | TRT2  | 0.037| 0.274| 0.276 | 94.9   | 94.3 | 1.091|
|        | MMRM-I| TRT1 | 0.005| 0.235| 0.266 | 97.6 | 2.4  | 1.112|
|        | TRT2  | 0.002| 0.306| 0.370 | 97.8   | 82.5 | 1.211|
|        | MMRM-II| TRT1| 0.012| 0.217| 0.213 | 94.4 | 5.6  | 1.028|
|        | TRT2  | 0.017| 0.274| 0.285 | 95.5   | 94.7 | 1.086|
|        | IMMRM | TRT1 | -0.001| 0.211| 0.200 | 93.5 | 6.5  | 1.000|
|        | TRT2  | 0.003| 0.252| 0.239 | 93.6   | 98.5 | 1.000|
Table 2: Simulation results comparing candidate estimators with 200 samples per arm under MCAR and MAR. For each estimator, we estimate the average treatment effect of TRT1 and TRT2, both comparing the control group. The following measures are used: bias, empirical standard error (ESE), averaged standard error (ASE), coverage probability (CP), probability of rejecting the null (PoR), relative mean squared error compared to IMMRM (RMSE). For RMSE, a number bigger than 1 indicates a larger mean squared error than IMMRM.

| Group | Bias  | ESE   | ASE   | CP(%) | PoR(%) | RMSE |
|-------|-------|-------|-------|-------|--------|------|
| **MCAR** |       |       |       |       |        |      |
| ANCOVA | TRT1  | 0.002 | 0.112 | 0.120 | 96.9   | 3.1  | 1.082 |
|        | TRT2  | 0.005 | 0.133 | 0.140 | 96.0   | 100.0| 1.091 |
| MMRM-I | TRT1  | 0.000 | 0.116 | 0.133 | 97.8   | 2.2  | 1.115 |
|        | TRT2  | 0.004 | 0.149 | 0.186 | 98.5   | 100.0| 1.224 |
| MMRM-II| TRT1  | 0.002 | 0.105 | 0.107 | 95.7   | 4.4  | 1.015 |
|        | TRT2  | 0.005 | 0.132 | 0.143 | 96.7   | 100.0| 1.082 |
| IMMRM  | TRT1  | 0.001 | 0.104 | 0.102 | 94.9   | 5.1  | 1.000 |
|        | TRT2  | 0.003 | 0.122 | 0.122 | 95.2   | 100.0| 1.000 |
| **MAR** |       |       |       |       |        |      |
| ANCOVA | TRT1  | 0.016 | 0.115 | 0.122 | 96.0   | 4.0  | 1.090 |
|        | TRT2  | 0.026 | 0.135 | 0.141 | 95.5   | 100.0| 1.089 |
| MMRM-I | TRT1  | 0.004 | 0.117 | 0.134 | 97.3   | 2.7  | 1.106 |
|        | TRT2  | 0.003 | 0.152 | 0.187 | 98.4   | 100.0| 1.221 |
| MMRM-II| TRT1  | 0.010 | 0.108 | 0.108 | 94.7   | 5.3  | 1.021 |
|        | TRT2  | 0.013 | 0.134 | 0.143 | 96.0   | 100.0| 1.082 |
| IMMRM  | TRT1  | 0.000 | 0.106 | 0.103 | 94.4   | 5.6  | 1.000 |
|        | TRT2  | 0.000 | 0.124 | 0.122 | 94.4   | 100.0| 1.000 |
7 Data application

We applied the ANCOVA, MMRM-I, MMRM-II, and IMMRM estimators to the IMAGINE-2 study. Due to very limited data for participants with no sulfonylurea/meglitinide use at baseline, we dropped this stratification variable in the analysis, resulting in 4 randomization strata in total. For each estimator, we computed the estimate, 95% confidence interval based on normal approximations, and the proportional variance reduction (PVR), defined as one minus the variance ratio of the estimator and the MMRM-I estimator. The standard error of each estimator is computed by the sandwich variance estimator as described in Section 5.

Table 3 summarizes our results of the data application. All estimators have similar treatment effect estimates, while their variances differ. Consistent with the results from our simulation study, IMMRM yields the most precise estimator, which is 2.0%, 4.7%, and 12.6% more efficient than MMRM-II, ANCOVA, and MMRM-I, respectively, indicating its advantage in increasing power.

Table 3: Summary of data application.

| Method     | Treatment effect estimate | 95% confidence interval | Proportional variance reduction compared to MMRM-I |
|------------|---------------------------|-------------------------|--------------------------------------------------|
| ANCOVA     | -0.283                    | (-0.380, -0.186)        | 8.3%                                             |
| MMRM-I     | -0.296                    | (-0.398, -0.194)        | 0.0%                                             |
| MMRM-II    | -0.299                    | (-0.395, -0.203)        | 10.8%                                            |
| IMMRM      | -0.292                    | (-0.388, -0.196)        | 12.6%                                            |

8 Discussion

For the analysis of longitudinal repeated measures data, we propose the IMMRM estimator that can improve precision for estimating the average treatment effect of the final outcome,
compared to standard practice that uses ANCOVA, MMRM-I, or MMRM-II. Such precision gain comes from appropriately adjusting for intermediate outcomes and accounting for stratified randomization, which can be translated into sample size reduction in trial planning. This result also applies to estimating the expectation of the final outcome for each treatment group.

Like most regression models, the price paid for precision gain is to estimate more parameters. IMMRM involves $K_p(J + 1) - p$ more parameters than MMRM-I and $K_pJ$ more parameters than MMRM-II. Given that most randomized clinical trials involve a few treatment groups and visits relative to the sample size (Mallinckrodt et al., 2008), the number of extra parameters is often dominated by the number of covariates. When the sample size is small, we recommend prespecifying a small number of prognostic baseline variables to avoid losing degrees of freedom.

Under the MAR assumption, an alternative approach for estimating the average treatment effect is the targeted minimum loss-based estimation (TMLE, van der Laan and Gruber, 2012). This method involves recursively fitting regression models for outcomes and propensity scores for non-missingness, and is consistent as long as one of the two sets of models is correctly specified. It is an open question, to the best of our knowledge, whether TMLE guarantees an efficiency gain by adjusting for intermediate outcomes while retaining its double-robustness.

Leveraging post-randomization information to improve precision is not limited to intermediate outcomes. Our main results also apply to adjustment for other post-randomization continuous-valued random variables measured before the final outcomes, such as the body mass index measured at each visit. When these additional variables provide new prognostic information beyond intermediate outcomes and covariates, adding them to the IMMRM model can bring further asymptotic precision gain.
Supplementary materials

The Supplementary material contains regularity conditions for our theorems, formulas of the asymptotic variances in the theorem and their consistent estimators, proofs, an example which shows MMRM-II is less precise than ANCOVA, and the missing data distribution under MAR in the simulation study.

References

Bugni, F. A., I. A. Canay, and A. M. Shaikh (2019). Inference under covariate-adaptive randomization with multiple treatments. *Quantitative Economics* 10(4), 1747–1785.

Davies, M., D. Russell-Jones, J.-L. Selam, T. Bailey, Z. Kerényi, J. Luo, J. Bue-Valleskey, T. Iványi, M. Hartman, J. Jacobson, et al. (2016). Basal insulin peglispro versus insulin glargine in insulin-naïve type 2 diabetes: Imagine 2 randomized trial. *Diabetes, Obesity and Metabolism* 18(11), 1055–1064.

Galbraith, S. and I. C. Marschner (2003). Interim analysis of continuous long-term endpoints in clinical trials with longitudinal outcomes. *Statistics in Medicine* 22(11), 1787–1805.

Gosho, M. and K. Maruo (2018). Effect of heteroscedasticity between treatment groups on mixed-effects models for repeated measures. *Pharmaceutical statistics* 17(5), 578–592.

Hampson, L. V. and C. Jennison (2013). Group sequential tests for delayed responses (with discussion). *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 75(1), 3–54.

Jiang, F., L. Tian, H. Fu, T. Hasegawa, and L. J. Wei (2018). Robust alternatives to ANCOVA for estimating the treatment effect via a randomized comparative study. *Journal of the American Statistical Association* 0, 1–37.
Kraemer, H. (2015). A source of false findings in published research studies: Adjusting for covariates. *JAMA Psychiatry* 72(10), 961–962.

Kunz, C. U., T. Friede, N. Parsons, S. Todd, and N. Stallard (2015). A comparison of methods for treatment selection in seamless phase II/III clinical trials incorporating information on short-term endpoints. *Journal of biopharmaceutical statistics* 25(1), 170–189.

Landau, S., R. Emsley, and G. Dunn (2018). Beyond total treatment effects in randomised controlled trials: Baseline measurement of intermediate outcomes needed to reduce confounding in mediation investigations. *Clinical Trials* 15(3), 247–256.

Lin, Y., M. Zhu, and Z. Su (2015). The pursuit of balance: An overview of covariate-adaptive randomization techniques in clinical trials. *Contemporary Clinical Trials* 45, 21 – 25. 10th Anniversary Special Issue.

Lu, X. and A. A. Tsiatis (2011, Oct). Semiparametric estimation of treatment effect with time-lagged response in the presence of informative censoring. *Lifetime Data Analysis* 17(4), 566–593.

Mallinckrodt, C., J. Bell, G. Liu, B. Ratitch, M. O’Kelly, I. Lipkovich, P. Singh, L. Xu, and G. Molenberghs (2020). Aligning estimators with estimands in clinical trials: putting the ich e9 (r1) guidelines into practice. *Therapeutic innovation & regulatory science* 54(2), 353–364.

Mallinckrodt, C. H., P. W. Lane, D. Schnell, Y. Peng, and J. P. Mancuso (2008). Recommendations for the primary analysis of continuous endpoints in longitudinal clinical trials. *Drug Information Journal* 42(4), 303–319.

Marschner, I. C. and S. L. Becker (2001). Interim monitoring of clinical trials based on long-term binary endpoints. *Statistics in Medicine* 20(2), 177–192.
McCulloch, C. E. and J. M. Neuhaus (2011). Misspecifying the shape of a random effects distribution: why getting it wrong may not matter. *Statistical science* 26(3), 388–402.

Neyman, J. S., D. M. Dabrowska, and T. Speed (1990). On the application of probability theory to agricultural experiments. Essay on principles. Section 9. *Statistical Science*, 465–472.

O’Kelly, M. and B. Ratitch (2014). *Clinical trials with missing data: a guide for practitioners*. John Wiley & Sons.

Qian, T., M. Rosenblum, and H. Qiu (2019). Improving precision through adjustment for prognostic variables in group sequential trial designs: Impact of baseline variables, short-term outcomes, and treatment effect heterogeneity. *arXiv:1910.05800*.

Qu, Y. and J. Luo (2015). Estimation of group means when adjusting for covariates in generalized linear models. *Pharmaceutical statistics* 14(1), 56–62.

Rubin, D. and M. van der Laan (2008). Covariate adjustment for the intention-to-treat parameter with empirical efficiency maximization. *U.C. Berkeley Division of Biostatistics Working Paper Series. Working Paper 229*.

Schuler, A. (2022). Mixed models for repeated measures should include time-by-covariate interactions to assure power gains and robustness against dropout bias relative to complete-case ANCOVA. *Therapeutic Innovation & Regulatory Science* 56(1), 145–154.

Shih, W. J. and H. Quan (1999). Planning and analysis of repeated measures at key time-points in clinical trials sponsored by pharmaceutical companies. *Statistics in Medicine* 18(8), 961–973.

Stallard, N. (2010). A confirmatory seamless phase ii/iii clinical trial design incorporating short-term endpoint information. *Statistics in Medicine* 29(9), 959–971.
Tsiatis, A. (2007). *Semiparametric theory and missing data*. Springer Science & Business Media.

van der Laan, M. J. and S. Gruber (2012). Targeted minimum loss based estimation of causal effects of multiple time point interventions. *The international journal of biostatistics* 8(1), article 9.

van der Vaart, A. (1998). *Asymptotic Statistics*. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press.

Van Lancker, K., A. Vandebosch, and S. Vansteelandt (2020). Improving interim decisions in randomized trials by exploiting information on short-term endpoints and prognostic baseline covariates. *Pharmaceutical statistics* 19(5), 583–601.

Wang, B., E. L. Ogburn, and M. Rosenblum (2019). Analysis of covariance in randomized trials: More precision and valid confidence intervals, without model assumptions. *Biometrics* 75(4), 1391–1400.

Wang, B., R. Susukida, R. Mojtabai, M. Amin-Esmaeili, and M. Rosenblum (2021). Model-robust inference for clinical trials that improve precision by stratified randomization and covariate adjustment. *Journal of the American Statistical Association*, 1–12.

Yang, L. and A. Tsiatis (2001). Efficiency study of estimators for a treatment effect in a pretest-posttest trial. *The American Statistician* 55(4), 314–321.

Ye, T., J. Shao, Y. Yao, and Q. Zhao (2021). Toward better practice of covariate adjustment in analyzing randomized clinical trials. *arXiv preprint arXiv:2009.11828v2*.

Zelen, M. (1974). The randomization and stratification of patients to clinical trials. *Journal of Chronic Diseases* 27(7), 365 – 375.
Zhou, M., Q. Tang, L. Lang, J. Xing, and K. Tatsuoka (2018). Predictive probability methods for interim monitoring in clinical trials with longitudinal outcomes. *Statistics in Medicine 37*(14), 2187–2207.
Supplementary Material to
Robustly leveraging post-randomization information to improve precision in randomized trials

Bingkai Wang\textsuperscript{1} and Yu Du\textsuperscript{2}

\textsuperscript{1}The Statistics and Data Science Department of the Wharton School, University of Pennsylvania, PA, USA
\textsuperscript{2}Statistics, Data and Analytics, Eli Lilly and Company, IN, USA

In Section A, we give some useful notations. In Section B, we provide the regularity conditions for Theorem 1 of the main paper. In Section C, we define the variance estimators for $\tilde{V}^{(est)}$ and $V^{(est)}$ and provide consistent estimators. In Section D, we introduce a few lemmas for proving our main results. In Section E, we prove Theorem 1, Corollary 1 and Corollary 2 of the main paper. In Section F, we give an example where MMRM-II is less precise than ANCOVA. In Section G, we provide the missing data mechanism for MAR in the simulation study.

A Notations

Let $\mathbf{1}_K$ be the column vector of length $K$ with each component equal to 1, $\mathbf{0}_K$ be the column vector of length $K$ with each component equal to 0, and $\mathbf{e}_t$ be the column vector of length $K$ with the $t$-th entry 1 and the rest 0. Let $\mathbf{I}_K$ be the $K \times K$ identity matrix. Let $\otimes$ be the Kronecker product. Let $I$ be the indicator function, i.e. $I\{A\} = 1$ if event $A$ is true and
0 otherwise. For any random vector \( \mathbf{W} \) with finite second-order moment, we define \( \tilde{\mathbf{W}} = \mathbf{W} - E[\mathbf{W}] \) and \( Var(\mathbf{W}) = E[\tilde{\mathbf{W}}\tilde{\mathbf{W}}^\top] \). For two random vectors \( \mathbf{W}_1 \) and \( \mathbf{W}_2 \), we define \( Cov(\mathbf{W}_1, \mathbf{W}_2) = E[\tilde{\mathbf{W}}_1\tilde{\mathbf{W}}_2^\top] \). Let \( \{0, 1\}^K \) be the set of \( K \)-dimensional binary vectors, i.e. \( \{0, 1\}^K = \{(x_1, \ldots, x_K)^\top : x_t \in \{0, 1\}, t = 1, \ldots K\} \). For any matrix (or vector), we use \( \| \cdot \| \) to denote its \( L_2 \) matrix (or vector) norm. For any vector \( \mathbf{v} = (v_1, \ldots, v_L) \), we use \( diag\{\mathbf{v}\} \) or \( \text{diag}\{v_l : l = 1, \ldots, L\} \) to denote an \( L \times L \) diagonal matrix with the diagonal entries being \( (v_1, \ldots, v_L) \). For any sequence \( x_1, \ldots, x_n, \ldots \), we define \( P_n x = n^{-1} \sum_{i=1}^n x_i \).

### B Regularity conditions

The regularity conditions for Theorem 1 are assumed on estimating equations \( \psi^{(\text{ANCOVA})}, \psi^{(\text{MMRM})} \) and \( \psi^{(\text{IMMRM})} \), which are defined by Equations (1), (3), and (5) below, respectively. Each of \( \psi^{(\text{ANCOVA})}, \psi^{(\text{MMRM})} \) and \( \psi^{(\text{IMMRM})} \) is a \( q \)-dimensional function of random variables \( (A, \mathbf{X}, \mathbf{Y}, M) \) and a set of parameters \( \theta \in \mathbb{R}^q \), where \( q \) and \( \theta \) vary among estimators, and \( \Delta \) are embedded in \( \theta \). Without causing confusion, we use \( \psi(A, \mathbf{X}, \mathbf{Y}, M; \theta) \) to represent any of the above three estimating equations. As we show in the proof of Theorem 1 below, each of the ANCOVA, MMRM-I, MMRM-II and IMMRM estimators is an \( \mathcal{M} \)-estimator, which is defined as the solution of \( \Delta \) to the equations \( P_n \psi(A, \mathbf{X}, \mathbf{Y}, M; \theta) = 0 \).

The regularity conditions are similar to those used in Section 5.3 of van der Vaart (1998) in their theorem on estimating equations \( \psi \) for showing asymptotic normality of \( \mathcal{M} \)-estimators for independent, identically distributed data. The regularity conditions are given below:

1. \( \theta \in \Theta \), a compact set in \( \mathbb{R}^q \).
2. \( E[||\psi(j, \mathbf{X}, \mathbf{Y}(j), M(j); \theta)||^2] < \infty \) for any \( \theta \in \Theta \) and \( j \in \{0, \ldots, J\} \).
(3) There exists a unique solution in the interior of Θ, denoted as θ, to the equations
\[ \sum_{j=0}^{J} \pi_j E[\psi(j, X, Y(j), M(j); \theta)] = 0. \]

(4) For each \( j \in \{0, \ldots, J\} \), the function \( \theta \mapsto \psi(j, x, y, m; \theta) \) is twice continuously differentiable for every \( (x, y, m) \) in the support of \( (X, Y(j), M(j)) \) and is dominated by an integrable function \( u(x, y, m) \).

(5) There exist a \( C > 0 \) and an integrable function \( v(X, Y(j), M(j)) \), such that, for each entry \( \psi_r, r = 1, \ldots, q \), of \( \psi \), \( ||\frac{\partial^2}{\partial \theta^2} \psi_r(j, x, y, m; \theta)|| < v(x, y, m) \) for every \( (j, x, y, m) \) in the support of \( (A, X, Y(j), M(j)) \) and \( \theta \) with \( ||\theta - \tilde{\theta}|| < C \).

(6) \( E \left[ \left| \frac{\partial}{\partial \theta} \psi(j, X, Y(j), M(j); \theta) \right|_{\theta=\hat{\theta}} \right|^2 \right] < \infty \) for \( j \in \{0, \ldots, J\} \) and
\[ \sum_{j=0}^{J} \pi_j E \left[ \frac{\partial}{\partial \theta} \psi(j, X, Y(j), M(j); \theta) \right|_{\theta=\hat{\theta}}^{-1} \] is invertible.

C Variance estimators in Theorem 1

For an M-estimator \( \hat{\theta} \in \mathbb{R}^q \) defined by \( P_n \psi(A, X, Y, M; \theta) = 0 \), its sandwich variance estimator under simple randomization is defined as
\[ \tilde{V}_n(\psi, \hat{\theta}) = \frac{1}{n} \left\{ P_n \left[ \frac{\partial}{\partial \theta} \psi(A, X, Y, M; \hat{\theta}) \right|_{\theta=\hat{\theta}} \right\}^{-1} \left\{ P_n \left[ \psi(A, X, Y, M; \theta) \psi(A, X, Y, M; \hat{\theta})^t \right] \right\} \left\{ P_n \left[ \frac{\partial}{\partial \theta} \psi(A, X, Y, M; \theta) \right|_{\theta=\hat{\theta}} \right\}^{-1,t}. \]

Since \( \Delta \) is embedded in \( \theta \), we can find \( C \in \mathbb{R}^{J \times q} \) such that \( \Delta = C\theta \). Then the variance estimator of \( \hat{\Delta} \) under simple randomization is defined as \( C \tilde{V}_n(\psi, \hat{\theta}) C^\top \).
For each est ∈ \{ANCOVA, MMRM-I, MMRM-II, IMMRRM\}, the variance estimator \( \tilde{V}_n^{(est)} \) is calculated by \( \mathbf{C} \tilde{V}_n(\psi^{(est)}, \hat{\theta}) \mathbf{C}^\top \). We note that the MMRM-I estimator and MMRM-II estimator share the same estimating equations \( \psi^{(MMRM)} \) with different specifications of \( u(X) \) as described in Equation (2).

We next define \( V_n^{(est)} \) for \( est = ANCOVA, MMRM-I, MMRM-II \). Define

\[
\tilde{V}_n(\psi^{(ANCOVA)}) = \sum_{s \in S} \frac{(P_n I\{S = s\} \mathbf{X})(P_n I\{S = s\} \mathbf{X})^\top}{P_n I\{S = s\}} - (P_n \mathbf{X})(P_n \mathbf{X})^\top,
\]

\[
\tilde{V}_n(\psi^{(MMRM-I)}) = \sum_{s \in S} \frac{(P_n I\{S = s\} \mathbf{X})(P_n I\{S = s\} \mathbf{X})^\top}{P_n I\{S = s\}} - (P_n \mathbf{X})(P_n \mathbf{X})^\top,
\]

\[
\tilde{V}_n(\psi^{(MMRM-II)}) = \sum_{s \in S} \frac{(P_n I\{S = s\} \mathbf{X})(P_n I\{S = s\} \mathbf{X})^\top}{P_n I\{S = s\}} - (P_n \mathbf{X})(P_n \mathbf{X})^\top,
\]

where \( \hat{\beta}_X \) is the MLE of \( \beta_X \) in the MMRM-I working model (2).

Following Equation (16), we define

\[
\mathbf{V}_n^{(ANCOVA)} = \tilde{V}_n^{(ANCOVA)} - \frac{1}{n} \mathbf{L}[\text{diag}\{\pi_j^{-1}(\hat{\mathbf{b}}_{Kj} - \hat{\mathbf{b}}_K)^\top \tilde{V}_n(\hat{\mathbf{E}}[\mathbf{X}|S])(\hat{\mathbf{b}}_{Kj} - \hat{\mathbf{b}}_K) : j = 0, \ldots, J\}]
\]

\[
- \hat{\mathbf{z}}^\top \tilde{V}_n(\hat{\mathbf{E}}[\mathbf{X}|S]) \hat{\mathbf{z}} \mathbf{L}^\top.
\]

Following Equation (15), we define

\[
\mathbf{V}_n^{(MMRM-I)} = \tilde{V}_n^{(MMRM-I)} - \frac{1}{n} \mathbf{L}[\text{diag}\{\pi_j^{-1}(\hat{\mathbf{b}}_{Kj} - \hat{\beta}_X)^\top \tilde{V}_n(\hat{\mathbf{E}}[\mathbf{X}|S])(\hat{\mathbf{b}}_{Kj} - \hat{\beta}_X) : j = 0, \ldots, J\}]
\]

\[
- \hat{\mathbf{v}}^\top \tilde{V}_n(\hat{\mathbf{E}}[\mathbf{X}|S]) \hat{\mathbf{v}} \mathbf{L}^\top.
\]
Similarly, we define
\[ V_n^{(MMRM-II)} = \tilde{V}_n^{(MMRM-II)} - \frac{1}{n} L \{ \text{diag} \{ \pi_j^{-1} (\hat{b}_{Kj} - \hat{b}_K) \}^\top \bar{\text{Var}}(\hat{E}[X|S]) (\hat{b}_{Kj} - \hat{b}_K) : j = 0, \ldots, J \} \]
\[ - \hat{z}^\top \bar{\text{Var}}(\hat{E}[X|S]) \hat{z} L^\top. \]

**D Lemmas**

**Lemma 1.** Let \( \Sigma \in \mathbb{R}^{K \times K} \) be a positive definite matrix. For each \( m \in \{0, 1\}^K \setminus \{0_K\} \), let \( n_m = \sum_{t=1}^{K} m_t \) be the number of ones in \( m \) and \( t_{m,1} < \cdots < t_{m,n_m} \) denote the ordered list of locations of ones in \( m \), i.e., \( m_t = 1 \) if \( t \in \{t_{m,1}, \ldots, t_{m,n_m}\} \) and 0 otherwise. We define
\[ D_m = [e_{t_{m,1}}, \ldots, e_{t_{m,n_m}}] \in \mathbb{R}^{K \times n_m} \]
and
\[ V_m(\Sigma) = I\{m \in \{0, 1\}^K \setminus \{0_K\}\}D_m(D_m^\top \Sigma D_m)^{-1}D_m^\top \in \mathbb{R}^{K \times K}, \]

which are deterministic functions of \( m \) and \( \Sigma \).

Let \( M = (M_1, \ldots, M_K) \) be a \( K \)-dimensional binary random vector taking values in \( \{0, 1\}^K \). We assume that \( P(M = 1_K) > 0 \). Then the following statements hold.

1. \( E[V_M(\Sigma)] \) is well-defined and positive definite.
2. \( e_K^\top E[V_M(\Sigma)]^{-1} e_K \leq P(M_K = 1) e_K^\top \Sigma e_K \). The equality holds if and only if either of the following conditions holds: (i) \( K = 1 \) or (ii) \( P(M_t = 1, M_K = 0) \sigma_{t,K} = 0 \) for \( t = 1, \ldots, K - 1 \), where \( \sigma_{t,K} = e_t^\top \Sigma e_K \) is the \((t,K)\)-th entry of \( \Sigma \).
3. Let \( A \in \mathbb{R}^{K \times K} \) be a positive definite matrix. Then
\[ e_K^\top E[V_M(\Sigma)]^{-1} e_K \leq e_K^\top E[V_M(A)]^{-1} E[V_M(\Sigma)V_M(A)] E[V_M(A)]^{-1} e_K. \] The equality holds if and only if
\[ P(M = m)e_K^\top E[V_M(A)]^{-1}V_m(A) = P(M = m)e_K^\top E[V_M(\Sigma)]^{-1}V_m(\Sigma) \text{ for all } m \in \{0, 1\}^K. \]

(4) Let \( B \in \mathbb{R}^{K \times K} \) be a positive semi-definite matrix. Then
\[
e_K^\top B e_K \leq e_K^\top E[V_M(\Sigma)]^{-1}E[V_M(\Sigma)BV_M(\Sigma)]E[V_M(\Sigma)]^{-1}e_K. \text{ The equality holds if and only if } P(M = m)e_K^\top E[V_M(\Sigma)]^{-1}V_m(\Sigma)B \text{ does not vary across } m \in \{0, 1\}^K. \]

(5) Letting \( C \in \mathbb{R}^{K \times K} \) be a positive definite matrix such that \( C - \Sigma \) is positive semi-definite, then
\[
e_K^\top (E[V_M(C)]^{-1} - E[V_M(\Sigma)]^{-1})e_K \geq e_K^\top (C - \Sigma)e_K. \]

Proof. (1) By definition, for any \( m \in \{0, 1\}^K \setminus \{0_K\} \), \( D_m \) is full column rank. Since \( \Sigma \) is positive definite and \( n_i \leq K \), then \( D_m \Sigma D_m \) is positive definite and \( V_m = D_m (D_m \Sigma D_m)^{-1} D_m^\top \) is positive semi-definite. In particular, for the case \( m = 1_K \), \( V_m = \Sigma^{-1} \) and so is positive definite. Since \( E[V_M(\Sigma)] = \sum_{m \in \{0, 1\}^K \setminus \{0_K\}} V_m P(M = m) \) and we assume \( P(M = 1_K) > 0 \), then for any \( x \in \mathbb{R}^K \),
\[
x^\top E[V_M]x = x^\top V_{1_K}x P(M = 1_K) + \sum_{m \in \{0, 1\}^K \setminus \{0_K, 1_K\}} x^\top V_m x P(M = m) \\
\geq x^\top V_{1_K}x P(M = 1_K) \\
> 0,
\]
which implies that \( E[V_M] \) is positive definite and so invertible.

(2) Consider the case \( K = 1 \). we have that \( \Sigma \) reduces to a positive number \( \sigma \) and \( V_m(\sigma) = I\{m = 1\}\sigma^{-1} \). Then \( e_K^\top E[V_m(\sigma)]^{-1}e_K = \frac{1}{P(m = 1)}\sigma = \frac{1}{P(M_K = 1)} e_K^\top \Sigma e_K \). We next consider the case that \( K \geq 2 \). Define \( \Omega_K = \{m \in \{0, 1\}^K \setminus \{e_K\} : m_K = 1\} \) and \( \Omega_{-K} = \{m \in \} \)
\{0, 1\}^K \setminus \{0_K\} : m_K = 0\}. Then we have

\[D_m = \begin{pmatrix} \tilde{D}_m & 0_{K-1} \\ 0_{m \times m-1}^\top & 1 \end{pmatrix}\] if \(m \in \Omega_K\) and

\[D_m = \begin{pmatrix} \tilde{D}_m \\ 0_{m \times m-1}^\top & 1 \end{pmatrix}\] if \(m \in \Omega_{-K}\),

where \(\tilde{D}_m \in \mathbb{R}^{(K-1) \times (m-1)}\) is a matrix taking the first \(K-1\) rows and first \((\sum_{t=1}^K m_t-1)\) columns of \(D_m\) if \(m \in \Omega_K\) and \(\tilde{D}_m \in \mathbb{R}^{(K-1) \times (m-1)}\) is the first \(K-1\) rows of \(\tilde{D}_m\) if \(m \in \Omega_{-K}\). We further define \(\Sigma_{-K,-K} \in \mathbb{R}^{(K-1) \times (K-1)}\), \(\Sigma_{-K,K} \in \mathbb{R}^{K-1}\) and \(\sigma = e_K^\top \Sigma e_K\) such that

\[\Sigma = \begin{pmatrix} \Sigma_{-K,-K} & \Sigma_{-K,K} \\ \Sigma_{-K,K} & \sigma \end{pmatrix} \].

Using the formula of block matrix inversion, we can compute that, if \(m \in \Omega_K\), then

\[V_m(\Sigma) = \begin{pmatrix} A_m & -\sigma^{-1} A_m \Sigma_{-K,K} \\ -\sigma^{-1} \Sigma_{-K,K} A_m & \sigma^{-1} + \sigma^{-2} \Sigma_{-K,K} A_m \Sigma_{-K,K} \end{pmatrix},\]

and, if \(m \in \Omega_{-K}\), then

\[V_m(\Sigma) = \begin{pmatrix} A_m & 0_{K-1} \\ 0_{K-1}^\top & 1 \end{pmatrix},\]

where \(A_m = \tilde{D}_m \left\{ \tilde{D}_m^\top (\Sigma_{-K,-K} - \Sigma_{-K,K} \Sigma_{-K,K}^\top / \sigma) \tilde{D}_m \right\}^{-1} \tilde{D}_m^\top \in \mathbb{R}^{(K-1) \times (K-1)}\) if \(m \in \Omega_K\) and

\[A_m = \tilde{D}_m \left\{ \tilde{D}_m^\top \Sigma_{-K,-K} \tilde{D}_m \right\}^{-1} \tilde{D}_m^\top \in \mathbb{R}^{(K-1) \times (K-1)}\] if \(m \in \Omega_{-K}\). Here \(A_m\) is well defined for each \(m \in \Omega_K \cup \Omega_{-K}\) since \(\tilde{D}_m^\top \Sigma \tilde{D}_m\) is positive definite. In addition, if \(m = e_K\), then \(D_m = e_K\) and \(V_m = \sigma^{-1} e_K e_K^\top\). Hence

\[E[V_M(\Sigma)] = \begin{pmatrix} \sum_{m \in \Omega_K \cup \Omega_{-K}} p_m A_m \\ -\sum_{m \in \Omega_K} p_m \sigma^{-1} A_m \Sigma_{-K,K} \\ \sum_{m \in \Omega_K \cup \Omega_{-K}} p_m A_m \Sigma_{-K,K} + \sigma^{-2} \Sigma_{-K,K} A_m \Sigma_{-K,K} + p e_K \sigma^{-1} \end{pmatrix} \].
Since we have shown that $E[V_M(\Sigma)]$ is positive definite and $P(M_K = 1) = p_{e_K} + \sum_{m \in \Omega_K} p_m$, by using the formula of block matrix inversion again, we have

\[
(e_K^\top E[V_m(\Sigma)]^{-1} e_K)^{-1}
\]

\[
= P(M_K = 1)\sigma^{-1} + \sum_{m \in \Omega_K} p_m \sigma^{-2} \Sigma_{-K,K}^T A_m \Sigma_{-K,K}
\]

\[
- \left( \sum_{m \in \Omega_K} p_m \sigma^{-1} \Sigma_{-K,K}^T A_m \right) \left( \sum_{m \in \Omega_K \cup \Omega_{-K}} p_m A_m \right)^{-1} \left( \sum_{m \in \Omega_K} p_m \sigma^{-1} A_m \Sigma_{-K,K} \right)
\]

\[
= P(M_K = 1)\sigma^{-1} + \sigma^{-2} \Sigma_{-K,K}^T \{ G_{\Omega_K} - G_{\Omega_K} (G_{\Omega_K} + G_{\Omega_{-K}})^{-1} G_{\Omega_K} \} \Sigma_{-K,K},
\]

where $G_{\Omega_K} = \sum_{m \in \Omega_K} p_m A_m$ and $G_{\Omega_{-K}} = \sum_{m \in \Omega_{-K}} p_m A_m$.

For $G_{\Omega_K}$, we note that $1_K \in \Omega_K$ with $p_{1_K} > 0$ and $G_{1_K} = \{ \Sigma_{-K,K} - \Sigma_{-K,K} \Sigma_{-K,K}^T / \sigma \}^{-1}$ is positive definite. Hence $G_{\Omega_K}$ is positive definite. Furthermore, $G_{\Omega_{-K}} \succeq 0$ by definition. Then

\[
G_{\Omega_K}^{-1} - (G_{\Omega_K} + G_{\Omega_{-K}})^{-1}
\]

\[
= (G_{\Omega_K} + G_{\Omega_{-K}})^{-1} \{ (G_{\Omega_K} + G_{\Omega_{-K}}) G_{\Omega_K}^{-1} (G_{\Omega_K} + G_{\Omega_{-K}}) - (G_{\Omega_K} + G_{\Omega_{-K}}) \} (G_{\Omega_K} + G_{\Omega_{-K}})^{-1}
\]

\[
= (G_{\Omega_K} + G_{\Omega_{-K}})^{-1} (G_{\Omega_{-K}} + G_{\Omega_{-K}} G_{\Omega_K}^{-1} G_{\Omega_{-K}}) (G_{\Omega_K} + G_{\Omega_{-K}})^{-1}
\]

\[
\succeq 0.
\]

Hence

\[
(e_K^\top E[V_m(\Sigma)]^{-1} e_K)^{-1}
\]

\[
= P(M_K = 1)\sigma^{-1} + \sigma^{-2} \Sigma_{-K,K}^T G_{\Omega_K} \{ G_{\Omega_K}^{-1} - (G_{\Omega_K} + G_{\Omega_{-K}})^{-1} \} G_{\Omega_K} \Sigma_{-K,K}
\]

\[
\geq P(M_K = 1)\sigma^{-1},
\]

8
which completes the proof of $e_K^T E[V_M(\Sigma)]^{-1} e_K \leq \frac{1}{P(M_k = 1)} e_K^T \Sigma e_K$.

We next examine when $e_K^T E[V_M(\Sigma)]^{-1} e_K = \frac{1}{P(M_k = 1)} e_K^T \Sigma e_K$. Since $\Sigma^{-1}_{\Omega_k}$ is positive semi-definite, then the above derivation shows that the equality holds if and only if $\Sigma_{-K,K}^T G_{\Omega_k} (G_{\Omega_k} + G_{\Omega_{-k}})^{-1} G_{\Omega_{-k}} = 0$. Noting that

$$G_{\Omega_k} (G_{\Omega_k} + G_{\Omega_{-k}})^{-1} G_{\Omega_{-k}} = G_{\Omega_{-k}} - G_{\Omega_k} (G_{\Omega_k} + G_{\Omega_{-k}})^{-1} G_{\Omega_k}$$

is symmetric, the equality holds if and only if $\Sigma_{-K,K}^T G_{\Omega_{-k}} (G_{\Omega_k} + G_{\Omega_{-k}})^{-1} G_{\Omega_k} = 0$. Since $(G_{\Omega_k} + G_{\Omega_{-k}})^{-1} G_{\Omega_k}$ is positive definite, we get

$$e_K^T E[V_M(\Sigma)]^{-1} e_K = \frac{1}{P(M_k = 1)} e_K^T \Sigma e_K$$

$$\Leftrightarrow \Sigma_{-K,K}^T G_{\Omega_{-k}} (G_{\Omega_k} + G_{\Omega_{-k}})^{-1} G_{\Omega_k} = 0$$

$$\Leftrightarrow \Sigma_{-K,K}^T G_{\Omega_k} G_{\Omega_{-k}} = 0$$

$$\Leftrightarrow \sum_{m \in \Omega_{-k}} p_m \Sigma_{-K,K}^T \tilde{D}_m \tilde{D}_m^T (\tilde{D}_m^T \Sigma_{-K,K} \tilde{D}_m)^{-1} \tilde{D}_m^T \Sigma_{-K,K} = 0$$

$$\Leftrightarrow p_m \Sigma_{-K,K}^T \tilde{D}_m = 0 \text{ for each } m \in \Omega_{-k}$$

$$\Leftrightarrow P(M_l = 1, M_k = 0) \sigma_{l,K} = 0 \text{ for each } t = 1, \ldots, K - 1,$$

where $\sigma_{j,K} = e_j^T \Sigma e_K$, which completes the proof.

(3) Denote $c_M^T(A) = e_K^T E[V_M(A)]^{-1} V_M(A)$ and $c_M^T(\Sigma) = e_K^T E[V_M(\Sigma)]^{-1} V_M(\Sigma)$. We have the following derivation:

$$E[[c_M^T(A) - c_M^T(\Sigma)] \Sigma (c_M^T(A) - c_M^T(\Sigma))]$$

$$= E[c_M^T(A) \Sigma c_M^T(A)] - E[c_M^T(A) \Sigma c_M^T(\Sigma)] - E[c_M^T(\Sigma) \Sigma c_M^T(A)] + E[c_M^T(\Sigma) \Sigma c_M^T(\Sigma)]$$

$$= E[c_M^T(A) \Sigma c_M^T(A)] - e_K^T E[V_M(\Sigma)]^{-1} e_K.$$
where the last equation comes from the fact that
\[ V_M(A) \Sigma V_M(\Sigma) = D_m (D_m^\top A D_m)^{-1} D_m^\top \Sigma D_m (D_m^\top \Sigma D_m)^{-1} D_m^\top = V_M(A) \]
\[ V_M(\Sigma) \Sigma V_M(\Sigma) = D_m (D_m^\top \Sigma D_m)^{-1} D_m^\top \Sigma D_m (D_m^\top \Sigma D_m)^{-1} D_m^\top = V_M(\Sigma). \]

Since \( \Sigma \) is positive definite, then we have \( E[c_M^\top (A) \Sigma c_M(A)] \geq e_K^\top E[V_M(\Sigma)]^{-1} e_K \), which is the desired inequality. The equality holds if and only if \( p_m \{ c_m(A) - c_M^\top (\Sigma) \} = 0 \) for all \( m \in \{0, 1\}^K \setminus \{0_K\} \).

(4) Define \( x_M^\top = e_K^\top E[V_M(\Sigma)]^{-1} V_M(\Sigma) B^\perp \). Here \( B^\perp \) is well-defined since \( B \) is positive semi-definite. Then we have
\[ e_K^\top E[V_M(\Sigma)]^{-1} E[V_M(\Sigma) B V_M(\Sigma)] E[V_M(\Sigma)]^{-1} e_K - e_K^\top B e_K \]
\[ = E[x_M^\top x_M] - E[x_M]^\top E[x_M] \]
\[ = \sum_{t=1}^{K} \text{Var}(x_M^\top e_t) \]
\[ \geq 0. \]

The equality holds if and only if \( \text{Var}(x_M^\top e_t) = 0 \) for \( t = 1, \ldots, K \), which is equivalent to \( x_M^\top \) being a constant vector.

(5) Define \( B = C - \Sigma \). The statement is proved by the following derivation:
\[ e_K^\top E[V_M(C)]^{-1} e_K \]
\[ = e_K^\top E[V_M(C)]^{-1} E[V_M(C) C V_M(C)] E[V_M(C)]^{-1} e_K \]
\[ = e_K^\top E[V_M(C)]^{-1} E[V_M(C) (B + \Sigma) V_M(C)] E[V_M(C)]^{-1} e_K \]
\[ \geq e_K^\top E[V_M(C)]^{-1} E[V_M(C) B V_M(C)] E[V_M(C)]^{-1} e_K + e_K^\top E[V_M(\Sigma)]^{-1} e_K \]
\[ \geq e_K^\top B e_K + e_K^\top E[V_M(\Sigma)]^{-1} e_K. \]
where the first inequality results from Lemma 1 (3), and the second inequality comes from Lemma 1 (4).

Lemma 2 (Kronecker product). Let $A \in \mathbb{R}^{n_1 \times n_2}, B \in \mathbb{R}^{n_3 \times n_4}, C \in \mathbb{R}^{n_2 \times n_5}, D \in \mathbb{R}^{n_4 \times n_6}$ be random matrices. Then

1. $(A \otimes B)(C \otimes D) = (AC) \otimes (BD)$,

2. If $A$ is independent of $(B, C)$, then $E[A \otimes B] = E[A] \otimes E[B]$ and $E[(AC) \otimes B] = E[(E[A]C) \otimes B]$.

3. $(A \otimes B)^{-1} = A^{-1} \otimes B^{-1}$ if $A$ and $B$ are invertible.

4. $(A \otimes B)^\top = A^\top \otimes B^\top$.

5. Suppose $n_1 = n_2, n_3 = n_4$, $A$ has eigenvalues $\lambda_1, \ldots, \lambda_{n_1}$, and $B$ has eigenvalues $\mu_1, \ldots, \mu_{n_3}$. Then $A \otimes B$ has eigenvalues $\lambda_i\mu_j$ for each $i = 1, \ldots, n_1$ and $j = 1, \ldots, n_3$.

Lemma 3. Given Assumption 1, for each $j = 0, \ldots, J$, let $Z_i(j) = h_j(Y_i(j), M_i(j), X_i) \in \mathbb{R}^q$ for some function $h_j$ such that $E[||Z_i(j)Z_i(j)^\top||] < \infty$. Then under stratified randomization,

$$
\frac{1}{\sqrt{n}} \sum_{i=1}^n \sum_{j=0}^J \left( I\{A_i = j\}Z_i(j) - \pi_jE[Z(j)] \right) \xrightarrow{d} N(0, G),
$$

where

$$
G = \sum_{j=0}^J \pi_jE[\text{Var}\{Z(j)|S]\} + \text{Var} \left( \sum_{j=0}^J \pi_jE[Z(j)|S]\right).
$$

Furthermore,

$$
\sum_{j=0}^J \pi_jE[Z(j)Z(j)^\top] - E \left[ \sum_{j=0}^J \pi_jZ(j) \right] E \left[ \sum_{j=0}^J \pi_jZ(j) \right]^\top - G = E[U(diag\{\pi\} - \pi\pi^\top)U^\top]
$$
is positive semi-definite, where

\[
U = (E[Z(0)|S], \ldots, E[Z(J)|S])
\]

\[
\pi = (\pi_0, \ldots, \pi_J)^\top.
\]

**Proof.** Let \( S = \{1, \ldots, R\} \) denote the levels in \( S \). Using the fact that \( E[Z_i(j)|S = S_i] = \sum_{s \in S} I\{S_i = s\} E[Z(j)|S = s] \) and \( E[Z_i(j)] = \sum_{s \in S} P(S = s) E[Z(j)|S = s] \), we have

\[
\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \sum_{j=0}^{J} \left( I\{A_i = j\} Z_i(j) - \pi_j E[Z(j)] \right)
\]

\[
= \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \sum_{j=0}^{J} \sum_{s \in S} I\{A_i = j, S_i = s\} \left( Z_i(j) - E[Z(j)|S = S_i] \right)
\]

\[
+ \sum_{s \in S} \sqrt{n} \left( \sum_{i=1}^{n} I\{S_i = s\} \right) - P(S = s) \right) \sum_{j=0}^{J} \pi_j E[Z(j)|S = s]
\]

\[
+ \sum_{s \in S} \sum_{j=0}^{J} \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left( I\{A_i = j, S_i = s\} - \pi_j I\{S_i = s\} \right) E[Z(j)|S = s]
\]

\[
= (1_{(J+1)L} \otimes I_q)^\top \mathbb{I}_{n}^{(1)} + u^\top \mathbb{I}_{n}^{(2)} + v_n \mathbb{I}_{n}^{(3)},
\]

where

\[
\mathbb{I}_{n}^{(1)} = \left( \frac{1}{\sqrt{n}} \sum_{i=1}^{n} I\{A_i = j, S_i = s\} \{Z_i(j) - E[Z(j)|S = S_i]\} : (j, s) \in \{0, \ldots, J\} \times S \right),
\]

\[
\mathbb{I}_{n}^{(2)} = \left( \sqrt{n} \left\{ \sum_{i=1}^{n} I\{S_i = s\} \right\} - P(S = s) \right) : s \in S),
\]

\[
\mathbb{I}_{n}^{(3)} = \left( \sqrt{n} \left\{ \sum_{i=1}^{n} I\{A_i = j, S_i = s\} \right\} - \pi_j \right) : (j, s) \in \{0, \ldots, J\} \times S),
\]

\[
u = \left( \sum_{j=0}^{J} \pi_j E[Z(j)|S = 1], \ldots, \sum_{j=0}^{J} \pi_j E[Z(j)|S = R] \right)^\top,
\]

\[
v_n = \left( \frac{\sum_{i=1}^{n} I\{S_i = s\}}{n} E[Z(j)|S = s] : (j, s) \in \{0, \ldots, J\} \times S \right)^\top,
\]
where \((x_{js} : (j, s) \in \{0, \ldots, J\} \times S) = (x_{01}^\top, \ldots, x_{0R}^\top, \ldots, x_{j1}^\top, \ldots, x_{jR}^\top)^\top\) and \((x_{js} : s \in S) = (x_{j1}^\top, \ldots, x_{jR}^\top)^\top\) for any vectors \(x_{js} \in \mathbb{R}^q\).

We next show that
\[
\left( \begin{array}{c}
\mathbb{L}_n^{(1)} \\
\mathbb{L}_n^{(2)} \\
\mathbb{L}_n^{(3)}
\end{array} \right) \overset{d}{\to} N\left( \begin{array}{ccc}
0 & 0 & 0 \\
0 & \Sigma_1 & 0 \\
0 & 0 & 0
\end{array} \right),
\]
where
\[
\Sigma_1 = bdiag\{\pi_j P(S = s) \text{Var}\{Z(j) | S = s\} : (j, s) \in \{0, \ldots, J\} \times S\},
\]
\[
p_S = (P(S = 1), \ldots, P(S = R))^\top,
\]
where \(bdiag\{V_{js} : (j, s) \in \{0, \ldots, J\} \times S\}\) represents a block diagonal matrix with \(V_{js}\) being the \((s-1)R+j+1\)-th diagonal block. The proof can be found in Lemma C.1 and C.2 of Appendix C of Bugni et al. (2019). The only difference is that \(Z_i(j)\) is substituted for \(Y_i(j)\) and all the arguments still hold.

By the delta method, we have \((1_{(J+1)L} \otimes I_q)^\top \mathbb{L}_n^{(1)} + u^\top \mathbb{L}_n^{(2)} \overset{d}{\to} N(0, \mathbf{G})\) and \(v \overset{P}{\to} v\) with \(v = (P(S = s)E[Z(j) | S = s] : (j, s) \in \{0, \ldots, J\} \times S)\). Using Slutsky’s theorem twice, we get the desired asymptotic normal distribution.
Finally, we have the following derivation:

\[
\sum_{j=0}^{J} \pi_j E[Z(j)Z(j)^\top] - E \left[ \sum_{j=0}^{J} \pi_j Z(j) \right] E \left[ \sum_{j=0}^{J} \pi_j Z(j) \right]^\top - G
\]

\[
= \sum_{j=0}^{J} \pi_j E[Z(j)Z(j)^\top] - E \left[ \sum_{j=0}^{J} \pi_j Z(j) \right] E \left[ \sum_{j=0}^{J} \pi_j Z(j) \right]^\top
\]

\[
- \sum_{j=0}^{J} \pi_j E[Z(j)Z(j)^\top] + \sum_{j=0}^{J} \pi_j E[E[Z(j)\mid S]E[Z(j)\mid S]^\top]
\]

\[
- E \left[ \sum_{j=0}^{J} \pi_j Z(j) \mid S \right] E \left[ \sum_{j=0}^{J} \pi_j Z(j) \mid S \right]^\top + E \left[ \sum_{j=0}^{J} \pi_j Z(j) \right] E \left[ \sum_{j=0}^{J} \pi_j Z(j) \right]^\top
\]

\[
= \sum_{j=0}^{J} \pi_j E[E[Z(j)\mid S]E[Z(j)\mid S]^\top] - E \left[ \sum_{j=0}^{J} \pi_j Z(j) \mid S \right] E \left[ \sum_{j=0}^{J} \pi_j Z(j) \mid S \right]^\top
\]

\[
= E[U(diag\{\pi\} - \pi\pi^\top)U^\top].
\]

Since \(diag\{\pi\} \succeq \pi\pi^\top\), then we get \(E[U(diag\{\pi\} - \pi\pi^\top)U^\top]\) is positive semi-definite. \(\square\)

**Lemma 4.** Given Assumption 1, under simple or stratified randomization, each data vector \((A_i, Y_i, M_i, X_i)\) is identically distributed and, for \(i = 1, \ldots, n\), \(A_i\) is independent of \(W_i\) and \(P(A_i = j) = \pi_j\).

Let \(P^*\) denote the distribution of \((A_i, Y_i, M_i, X_i)\) and \(E^*\) be the associated expectation. Define \(Z = f(Y, M, X)\) and \(Z(j) = f(Y(j), M(j), X)\) such that \(E[Z(j)^2] < \infty\) for \(j = 0, \ldots, J\). Then \(E^*[I\{A = j\}Z] = \pi_j E[Z(j)]\) and \(E^*[I\{A = j\}Z\mid S] = \pi_j E[Z(j)\mid S]\) for \(j = 0, \ldots, J\).

**Proof.** See Lemma 4 and Lemma 3 in the Supplementary of Wang et al. (2021). The only difference of proof is that \(A = 1\) is substituted by \(A = j\) for \(j = 1, \ldots, J\), and \((Y, M)\) are substituted for \((Y, M)\). \(\square\)
E Proofs

E.1 Proof of Theorem 1

Outline of the proof: Consider the estimator \( \hat{\Delta}^{(\text{est})} \) for each est \( \in \{\text{ANCOVA, MMRM-I, MMRM-II, IMMRM}\} \). We first show that \( \Delta^{(\text{est})} \) is an M-estimator. We then apply Theorem 1 of Wang et al. (2021) to show that \( \Delta^{(\text{est})} \) is model-robust and asymptotically linear with influence function \( IF^{(\text{est})} \). The influence function \( IF^{(\text{est})} \) is the same under simple and stratified randomization. Next, we prove the asymptotic normality by Lemma 3, which is a central limit theorem for sums of random vectors under stratified randomization that generalizes Lemma B.2 of Bugni et al. (2019). Next, we calculate \( IF^{(\text{est})} \) and derive the asymptotic covariance matrix i.e., \( V^{(\text{est})} \) and \( \tilde{V}^{(\text{est})} \), for which the detailed algebra is given in Lemma 5 and Lemma 6. Finally, we compare the asymptotic covariance matrices, where Lemma 1 is used to handle missing data.

Proof of Theorem 1. The ANCOVA estimator can be computed by solving

\[
P_n \psi^{(\text{ANCOVA})}(A, X, Y, M; \theta) = 0,
\]

where

\[
\psi^{(\text{ANCOVA})}(A, X, Y, M; \theta) = I\{M_K = 1\}(Y_K - \beta_0 - \sum_{j=1}^J \beta_{A_j}I\{A = j\} - \beta_X^T X)
\]

(1)

where \( A = (I\{A = 1\}, \ldots, I\{A = J\})^T \) is a vector of treatment assignment indicator and \( \theta = (\beta_0, \beta_{A_1}, \ldots, \beta_{A_J}, \beta_X^T)^T \). Hence \( \Delta^{(\text{ANCOVA})} \) is an M-estimator.

The MMRM-I and MMRM-II working models can be rewritten in one formula below:

\[
Y = \beta_0 + (I_K \otimes A)^T \beta_A + u(X)^T \beta_{u(X)} + \varepsilon,
\]

(2)
We further define $Q_i = \left(\beta_0, \ldots, \beta_0K\right)^\top \in \mathbb{R}^K$, $\beta_A = (\beta_{A11}, \ldots, \beta_{A11}, \ldots, \beta_{A1K}, \ldots, \beta_{A1K})^\top \in \mathbb{R}^{JK}$, $\beta_{u(X)} \in \mathbb{R}^q$ are column vectors of parameters, $u(X) \in \mathbb{R}^{q_u \times K}$ is a matrix function of $X$, and the error terms $\varepsilon \sim N(0, \Sigma)$, where $\Sigma \in \mathbb{R}^{K \times K}$ is a positive-definite covariance matrix. For MMRM-I, $q_u = p$ (the dimension of $X$), $u(X) = X1_K^\top$ and $\beta_{u(X)} = \beta_X$. For MMRM-II, $q = pK$, $u(X) = I_K \otimes X$ and $\beta_{u(X)} = (\beta_{X1}, \ldots, \beta_{XK})^\top$.

Under the working model, the random error vectors $\varepsilon_i$, $i = 1, \ldots, n$, are assumed to be independent (of each other and of $\{(A_i, X_i)\}_{i=1}^n$), identically distributed draws from $N(0, \Sigma)$. Denote $\Sigma = \Sigma(\alpha)$, where $\alpha = (\alpha_1, \ldots, \alpha_L)^\top \in \mathbb{R}^L$ is the vector of unknown parameters in $\Sigma$. For example, $\alpha$ consists of the lower triangular and diagonal entries of $\Sigma$ if no structure is assumed on $\Sigma$.

For each $i$, let $n_i = \sum_{t=1}^K M_{it}$ be the number of non-missing outcomes and $Y_i^o \in \mathbb{R}^{n_i}$ be the observed outcomes if $n_i > 0$. Let $t_{i,1} < \cdots < t_{i,n_i}$ denote the ordered list of visits when the outcomes are not missing. For example, $t_{i,1}$ is the first non-missing visit for subject $i$.

We define $D_{M_i} = [e_{t_{i,1}}, e_{t_{i,2}}, \ldots, e_{t_{i,n_i}}] \in \mathbb{R}^{K \times n_i}$. We use the subscript $M_i$ to note that $D_{M_i}$ is a deterministic function of $M_i$. Then $Y_i^o = D_{M_i}^\top Y_i$. The observed data vector for each $i$ is $(Y_i^o, M_i, A_i, X_i)$.

Denote the full set of parameters as $\theta = (\beta^\top, \alpha^\top)^\top$, where $\beta = (\beta_0^\top, \beta_A^\top, \beta_{u(X)}^\top)^\top$. We further define $Q_i = [I_K, (I_K \otimes A_i)^\top, u(X_i)^\top]^\top$. We let $Q_i$ denote $Q$ with $A_i, X_i$ substituted for $A, X$. It follows that $Q_i^\top \beta = \beta_0 + (I_K \otimes A_i)^\top \beta_A + u(X_i)^\top \beta_{u(X)}$. Then we have $Y_i^o | (A_i, X_i, M_i, M_i \neq 0_K) \sim N(D_{M_i}^\top Q_i^\top \beta, D_{M_i}^\top \Sigma D_{M_i})$ under the MMRM-I or MMRM-II working model assumptions and missing completely at random (MCAR).

The corresponding log likelihood function conditional on $\{(A_i, X_i, M_i)\}_{i=1}^n$ is a constant.
independent of the parameter vector $\theta$ plus the following:

$$-\frac{1}{2} \sum_{i=1}^{n} I\{M_i \neq 0_K\} \left\{ \log |D_M^\top \Sigma D_M| + (Y_i^o - D_M^\top Q_i^\top \beta)^\top (D_M^\top \Sigma D_M)^{-1} (Y_i^o - D_M^\top Q_i^\top \beta) \right\}$$

$$= -\frac{1}{2} \sum_{i=1}^{n} I\{M_i \neq 0_K\} \left\{ \log |D_M^\top \Sigma D_M| + (Y_i - Q_i^\top \beta)^\top D_M (D_M^\top \Sigma D_M)^{-1} D_M^\top (Y_i - Q_i^\top \beta) \right\}$$

$$= \frac{n}{2} P_n l(\theta; Y^o|A, X, M),$$

where we define

$$l(\theta; Y^o|A, X, M) = -I\{M \neq 0_K\} \left\{ \log |D_M^\top \Sigma D_M| + (Y - Q^\top \beta)^\top D_M (D_M^\top \Sigma D_M)^{-1} D_M^\top (Y - Q^\top \beta) \right\}.$$ 

To derive the estimating functions for the corresponding maximum likelihood estimator, we use the following results to compute the differential of $l(\theta; Y^o|A, X, M)$ with respect to $\theta$. By Equation (8.7) of Dwyer (1967), we have

$$\frac{\partial \log(|D_M^\top \Sigma D_M|)}{\partial \alpha_j} = \text{tr} \left( \frac{\partial \log(|D_M^\top \Sigma D_M|)}{\partial \Sigma} \frac{\partial \Sigma}{\partial \alpha_j} \right) = \text{tr} \left( D_M (D_M^\top \Sigma D_M)^{-1} D_M^\top \frac{\partial \Sigma}{\partial \alpha_j} \right).$$

By Theorem 5 of MacRae et al. (1974), we have

$$\frac{\partial (D_M^\top \Sigma D_M)^{-1}}{\partial \alpha_j} = - (D_M^\top \Sigma D_M)^{-1} D_M^\top \frac{\partial \Sigma}{\partial \alpha_j} D_M (D_M^\top \Sigma D_M)^{-1}.$$ 

Denoting $V_M(\Sigma) = I\{M \neq 0_K\} D_M (D_M^\top \Sigma D_M)^{-1} D_M^\top$, we have shown that

$$\frac{\partial V_M(\Sigma)}{\partial \alpha_j} = - V_M(\Sigma) \frac{\partial \Sigma}{\partial \alpha_j}. V_M(\Sigma).$$

Using the above results, the estimating functions for the MLE $\hat{\theta}$ for $\theta$ under MMRM-I
or MMRM-II are

\[
\psi^{(MMRM)}(A, X, Y, M; \theta) = \begin{pmatrix}
    QV_M(Y - Q^\top \beta) \\
    -\text{tr}(V_M \frac{\partial \Sigma}{\partial \alpha_l}) + (Y - Q^\top \beta)^\top V_M \frac{\partial \Sigma}{\partial \alpha_l} V_M(Y - Q^\top \beta),
\end{pmatrix},
\]

which implies that \(\widehat{\Delta}^{(MMRM-I)}\) and \(\widehat{\Delta}^{(MMRM-II)}\) are M-estimators. In the above expression of \(\psi^{(MMRM)}\), we omit \((\Sigma)\) from \(V_M(\Sigma)\) for conciseness. We note that \(V_M\) is a random matrix taking values in \(\mathbb{R}^{K \times K}\) and defined in the same way as in Lemma 1.

The IMMRM working model (3) can be written as

\[
Y = \beta_0 + (I_K \otimes A)^\top \beta_A + (I_K \otimes X)^\top \beta_{I_k \otimes X} + (I_K \otimes X \otimes A)^\top \beta_{AX} + \varepsilon_A,
\]

where \(\beta_0, \beta_A, A\) are defined in Equation (2), \(\beta_{I_k \otimes X} = (\beta_{X1}, \ldots, \beta_{XK})^\top, \beta_{AX} \in \mathbb{R}^{JpK}\) with the \(\{Jp(k-1) + J(m-1) + j\}\)-th entry being \(\beta_{AX,m,jk}\) for \(j = 1, \ldots, J, k = 1, \ldots, K\) and \(m = 1, \ldots, p\), and \(\varepsilon_A = \sum_{j=0}^J I\{A = j\} \varepsilon_j\), where \(\varepsilon_j \sim N(0, \Sigma_j)\) and \((\varepsilon_0, \ldots, \varepsilon_J)\) are independent of each other. Let \(\alpha_j \in \mathbb{R}^L\) be the unknown parameters in \(\Sigma_j\) for \(j = 0, \ldots, J\). We define \(\gamma = (\beta_0^\top, \beta_A^\top, \beta_{I_k \otimes X}^\top, \beta_{AX}^\top)^\top\) and \(\theta = (\Delta, \gamma^\top, \alpha_0^\top, \ldots, \alpha_J^\top)^\top\). Following a similar procedure to MMRM-I, the estimating functions for the MLE \(\hat{\theta}^{(IMMRM)}\) under IMMRM are

\[
\psi^{(IMMRM)}(A, X, Y, M; \theta) = \begin{pmatrix}
    \beta_{Ajk} + X^\top \beta_{AX,jk} - \Delta_j, \ j = 1, \ldots, J \\
    RV_{AM}(Y - R^\top \gamma) \\
    I\{A = j\} \left( -\text{tr}(V_{AM} \frac{\partial \Sigma}{\partial \alpha_l}) + (Y - R^\top \gamma)^\top V_{AM} \frac{\partial \Sigma}{\partial \alpha_l} V_{AM}(Y - R^\top \gamma) \right),
\end{pmatrix},
\]

where \(R = [I_K (I_K \otimes A)^\top (I_K \otimes X)^\top (I_K \otimes X \otimes A)^\top]^\top\) and

\[
V_{AM} = V_M(\sum_{j=0}^J I\{A = j\} \Sigma_j) = I\{M \neq 0_K\} D_M(\sum_{j=0}^J I\{A = j\} D_M^\top \Sigma_j D_M)^{-1} D_M^\top.
\]

Hence \(\widehat{\Delta}^{(IMMRM)}\), as the first \(J\) entries of \(\hat{\theta}^{(IMMRM)}\), is an M-estimator.
For each est ∈ \{ANCOVA, MMRM-I, MMRM-II, IMMRM\}, we have just shown that $\hat{\Delta}^{(est)}$ is an M-estimator. By Assumption 1 and regularity conditions, we apply Theorem 1 of Wang et al. (2021) and get, under simple or stratified randomization,

$$\sqrt{n}(\hat{\Delta}^{(est)} - \Delta^{(est)}) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} IF^{(est)}(A_i, X_i, Y_i, M_i) + o_p(1),$$  \tag{6}$$

where $\Delta^{(est)}$ satisfies $E^*[\psi^{(est)}(A, X, Y, M; \theta)] = 0$ with $E^*$ defined in Lemma 4, and $IF^{(est)}$ represents the $J$-dimensional influence function. We note that Theorem 1 of Wang et al. (2021) is developed for binary treatment (i.e. $J = 1$) and scalar outcome (i.e. $K = 1$), but their proof can be easily generalized to accommodate multiple treatment arms and repeated measured outcomes (as in Example 3 of Wang et al., 2021).

We next show $\hat{\Delta}^{(MMRM-I)}$ is model-robust. By $E^*[\psi^{(MMRM)}(A, X, Y, M; \theta)] = 0$, we have

$$E^*[V_M\{Y - \beta_0 - (I_K \otimes A)^\top \beta_{A} - u(X)^\top \beta_{u(X)}\}] = 0,$$

$$E^*[(I_K \otimes A)V_M\{Y - \beta_0 - (I_K \otimes A)^\top \beta_{A} - u(X)^\top \beta_{u(X)}\}] = 0,$$

which are first $2K$ equations in $E^*[\psi(A, X, Y, M; \theta)] = 0$, where $V_M = V_M(\Sigma^{(MMRM-I)})$ with $\Sigma^{(MMRM-I)}$ being the probability limit of $\Sigma(\hat{\alpha})$ under MMRM-I. By the MCAR assumption and Lemma 4, the first $K$ equations imply that

$$E^*[V_M]\sum_{j=0}^{J} \{E[Y(j)] - \beta_0 - \beta_{A_j} - E[u(X)^\top \beta_{u(X)}]\} \pi_j = 0,$$

where $\beta_{A_j} = (\beta_{A_{j1}}, \ldots, \beta_{A_{jK}})^\top$ for $j = 1, \ldots, J$ and $\beta_{A_j} = 0_K$. Similarly, the $(K + 1) - 2K$th equations imply that, for $j = 1, \ldots, K$ and $t = 1, \ldots, K$,

$$E^*[e_t^\top V_M]\{E[Y(j)] - \beta_0 - \beta_{A_j} - E[u(X)^\top \beta_{u(X)}]\} \pi_j = 0.$$
The above two sets of equations and the positivity assumption imply that, for \( j = 0, \ldots, J \),

\[
E^*\{V_M\}\{E[Y(j)] - \beta_j - \beta_{Aj} - E[u(X)^\top \beta_{u(x)}]\} = 0.
\]

The assumption \( P(M(j) = 1_K) > 0 \) and Lemma 1 (1) implies that \( E^*\{V_M\} \) is invertible. Then the above equations imply that, for \( j = 1, \ldots, J \),

\[
\beta_{Aj} = \beta_{Aj} - \beta_{A0} = E[Y(j)] - E[Y(0)]
\]

and hence \( \Delta^{(MMRM-I)} = (\beta_{A1K}, \ldots, \beta_{AJK})^\top = \Delta^* \). The above proof also applies to the MMRM-II estimator by substituting \( V_M(\Sigma^{(MMRM-II)}) \) for \( V_M \), which implies that \( \hat{\Delta}^{(MMRM-II)} \) is model-robust. Also, since the ANCOVA estimator is a special case of the MMRM-II estimator setting \( K = 1 \), we get that the ANCOVA estimator is model-robust.

Following a similar procedure, we next show that \( \hat{\Delta}^{(IMMRM)} \) is model-robust. We have that, for \( j = 1, \ldots, J \),

\[
E[V_{j,M}]\{E[Y(j)] - \beta_j - \beta_{Aj} - E[(I_K \otimes X)^\top \beta_{I_K \otimes X}] - E[(I_K \otimes X \otimes e_j)^\top \beta_{AX}]\} \pi_j = 0,
\]

and

\[
E[V_{0,M}]\{E[Y(0)] - \beta_j - E[(I_K \otimes X)^\top \beta_{I_K \otimes X}]\} \pi_0 = 0,
\]

where \( V_{j,M} = I\{M(j) \neq 0_K\}D_M(j)(D_M(j)^\top \Sigma_{j}^{(IMMRM)}D_M(j))^{-1}D_M(j)^\top \) with \( \Sigma_{j}^{(IMMRM)} \) being the probability limit of \( \Sigma_{j}(\hat{\alpha}_j) \) in the IMMRM model (3) for \( j = 0, \ldots, J \). The assumption \( P(M(j) = 1_K) > 0 \) and Lemma 1 (1) implies that \( E[V_{j,M}] \) is invertible. Thus, for \( j = 1, \ldots, J \),

\[
E[Y(j)] - E[Y(0)] = \beta_{Aj} + E[(I_K \otimes X \otimes e_j)^\top \beta_{AX}],
\]
which implies \(E[Y_K(j)] - E[Y_K(0)] = \beta_{AjK} + \beta_{AXjK}^\top X_j\). Since the first equation of \(\psi^{(IMMRM)}\) indicates that \(\beta_{AjK} + \beta_{AXjK}^\top X_j = \Delta_j\), we get \(\Delta_j = E[Y_K(j)] - E[Y_K(0)] = \Delta^*_j\), which completes the proof of model-robustness of \(\hat{\Delta}^{(IMMRM)}\).

We next prove that \(\sqrt{n}(\hat{\Delta}^{(est)} - \Delta^*)\) weakly converges to a normal distribution, under simple or stratified randomization. Given Equations (6), it suffices to show that \(\frac{1}{\sqrt{n}} \sum_{i=1}^{n} IF^{(est)}(A_i, X_i, Y_i, M_i)\) weakly converges to a normal distribution. Under simple randomization, \((A_i, X_i, Y_i, M_i), i = 1, \ldots, n\) are independent to each other and identically distributed. Since the regularity conditions implied that \(IF^{(est)}\) has finite second moment, the central limit theorem implies the desired weak convergence. Furthermore, we have \(\hat{V}^{(est)} = \text{Var}^*(IF^{(est)}(A, X, Y, M))\). Under stratified randomization, we define \(Z_i(j) = IF^{(est)}(j, X_i, Y_i(j), M_i(j))\) for \(j = 0, \ldots, J\). Then \(IF^{(est)}(A_i, X_i, Y_i, M_i) = \sum_{j=0}^J I\{A_i = j\} Z_i(j)\). Since \(E^*[IF^{(est)}(A, X, Y, M)] = 0\), Lemma 4 implies that \(\sum_{j=0}^J \pi_j E[Z_i(j)] = 0\).

By the regularity conditions, we apply Lemma 3 and get

\[
\frac{1}{\sqrt{n}} \sum_{i=1}^{n} IF^{(est)}(A_i, X_i, Y_i, M_i) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \sum_{j=0}^J (I\{A_i = j\} Z_i(j) - \pi_j E[Z_i(j)]) \xrightarrow{d} N(0, G),
\]

which completes the proof of asymptotic normality. In addition, Lemma 3 also implies that \(\hat{V}^{(est)} \succeq V^{(est)}\).

For the ANCOVA, MMRM-I, MMRM-II and IMMRM estimators, the influence functions by Lemmas 5 and 6 are given below:

\[
IF^{(ANCOVA)} = LT^{(ANCOVA)} \{Y - h(A, X)\},
\]

\[
IF^{(MMRM-I)} = LT^{(MMRM-I)}(Y - Q^\top \beta),
\]

\[
IF^{(MMRM-II)} = LT^{(MMRM-II)} \{Y - h(A, X)\},
\]

\[
IF^{(IMMRM)} = LT^{(IMMRM)}(Y - R^\top \gamma) + Lt^\top(X - E[X]),
\]
where

\[
L = (-1_J \quad I_J) \in \mathbb{R}^{J \times (J+1)},
\]

\[
T^{(\text{ANCOVA})} = \left( \frac{I\{A = 0\}}{\pi_0} V M E^* [V_M]^{-1} e_K, \ldots, \frac{I\{A = J\}}{\pi_J} V M E^* [V_M]^{-1} e_K \right)^T,
\]

\[
T^{(\text{MMRM-I})} = \left( \frac{I\{A = 0\}}{\pi_0} V M E^* [V_M]^{-1} e_K, \ldots, \frac{I\{A = J\}}{\pi_J} V M E^* [V_M]^{-1} e_K \right)^T,
\]

\[
T^{(\text{MMRM-II})} = \left( \frac{I\{A = 0\}}{\pi_0} V M E^* [V_M]^{-1} e_K, \ldots, \frac{I\{A = J\}}{\pi_J} V M E^* [V_M]^{-1} e_K \right)^T,
\]

\[
T^{(\text{IMMRM})} = \left( \frac{I\{A = 0\}}{\pi_0} V M E^* [V_M]^{-1} e_K, \ldots, \frac{I\{A = J\}}{\pi_J} V M E^* [V_M]^{-1} e_K \right)^T.
\]

\[
Y - h(A, X) = \sum_{j=0}^{J} I\{A = j\} \left\{ Y - E[Y(j)] - Cov^*(Y, X) Var(X)^{-1}(X - E[X]) \right\},
\]

\[
Y - Q^\top \beta = \sum_{j=1}^{J} I\{A = j\} \left\{ Y(j) - E[Y(j)] - 1_K(X - E[X])^\top \beta_X \right\},
\]

\[
Y - R^\top \gamma = \sum_{j=1}^{J} I\{A = j\} \left\{ Y(j) - E[Y(j)] - Cov(Y(j), X) Var(X)^{-1}(X - E[X]) \right\},
\]

\[
r = (b_{K0}, \ldots, b_{KJ}),
\]

where

\[
V_M = V_M (s_{(\text{MMRM-I})}^{(\text{MMRM-I})}) = V_M (E^* [(Y - Q^\top \beta) (Y - Q^\top \beta)^\top]),
\]

\[
\Sigma_M = V_M (s_{(\text{MMRM-II})}^{(\text{MMRM-II})}) = V_M (E^* [(Y - h(A, X)) (Y - h(A, X))^\top]),
\]

\[
\Sigma_{iM} = V_M (s_{i}^{(\text{IMMRM})}) = V_M \left( E^* \left[ \frac{I\{A = j\}}{\pi_j} (Y - R^\top \gamma)(Y - R^\top \gamma)^\top \right] \right), \quad j = 0, \ldots, J,
\]

\[
\beta_X = Var(X)^{-1} Cov^*(Y, X) \frac{E^* [V_M] 1_K}{1_K E^* [V_M] 1_K},
\]

\[
b_{Kj} = Var(X)^{-1} Cov(X, Y_K(j)), \quad j = 0, \ldots, J.
\]
Furthermore, we have

\[ \tilde{V}^{(\text{ANCOVA})} = \frac{1}{P^*(M_K = 1)} L \left( \text{diag}\{\pi_j^{-1} e_K^T \Sigma_j^{(\text{ANCOVA})} e_K : j = 0, \ldots, J\} \right) L^T, \]  

(11)

\[ \tilde{V}^{(\text{MMRM-I})} = L \text{diag} \left\{ e_K^T E^*[\Sigma_M]^{-1} E^* \left[ \pi_j^{-1} \Sigma_j^{(\text{MMRM-I})} \Sigma_M \right] E^*[\Sigma_M]^{-1} e_K : j = 0, \ldots, J \right\} L^T, \]  

(12)

\[ \tilde{V}^{(\text{MMRM-II})} = L \text{diag} \left\{ e_K^T E^*[\Sigma_M]^{-1} E^* \left[ \pi_j^{-1} \Sigma_j^{(\text{MMRM-II})} \Sigma_M \right] E^*[\Sigma_M]^{-1} e_K : j = 0, \ldots, J \right\} L^T, \]  

(13)

\[ \tilde{V}^{(\text{IMMRM})} = L \left( \text{diag}\{e_K^T E^*[\Sigma_M]^{-1} e_K : j = 0, \ldots, J\} + r^T \text{Var}(X)r \right) L^T, \]  

(14)

where

\[ \Sigma_j^{(\text{ANCOVA})} = \Sigma_j^{(\text{MMRM-II})} = E^* \left[ I \{ A = j \} \left\{ Y - h(A, X) \right\} \left\{ Y - h(A, X) \right\}^T \right], \]

\[ \Sigma_j^{(\text{MMRM-I})} = E^* \left[ I \{ A = j \} \left( Y - Q^\top \beta \right) \left( Y - Q^\top \beta \right)^T \right]. \]

We next compute \( V^{(\text{est})} \). Lemma 6 implies that, for \( j = 0, \ldots, J \), \( E[Y_K(j)|S] = E[Y_K(j)] + b_{Kj}^\top (E[X|S] - E[X]) \). Then using Equation (8), we get

\[ E[I\tilde{F}^{(\text{MMRM-I})}(j, X_i, Y_i(j), M_i(j))|S] = L \tilde{e}_{j+1} \pi_j^{-1} (b_{Kj} - \underline{\beta}_X)^\top (E[X|S] - E[X]), \]

where \( \tilde{e}_{j+1} \in \mathbb{R}^{J+1} \) has the \((j+1)\)-th entry 1 and the rest 0. Hence by Lemma 3, we have

\[ V^{(\text{MMRM-I})} = \tilde{V}^{(\text{MMRM-I})} - L [\text{diag}\{\pi_j^{-1} (b_{Kj} - \underline{\beta}_X)^\top \text{Var}(E[X|S]) (b_{Kj} - \underline{\beta}_X) : j = 0, \ldots, J\} - v^\top \text{Var}(E[X|S])v] L^T, \]  

(15)
where $v = (b_{K0} - \beta_x, \ldots, b_{KJ} - \beta_x)$. Similarly, we have $V^{(IMMRM)} = \tilde{V}^{(IMMRM)}$ and

$$V^{(ANCOVA)} = \tilde{V}^{(ANCOVA)} - L[diag\{\pi_j^{-1}(b_{Kj} - b_K)^\top Var(E[X|S])(b_{Kj} - b_K) : j = 0, \ldots, J\} - z^\top Var(E[X|S])z]L^\top,$$  

(16)

$$V^{(MMRM-II)} = \tilde{V}^{(MMRM-II)} - L[diag\{\pi_j^{-1}(b_{Kj} - b_K)^\top Var(E[X|S])(b_{Kj} - b_K) : j = 0, \ldots, J\} - zVar(E[X|S])z^\top]L^\top,$$

(17)

where $b_K = Var(X)^{-1}Cov*(X, Y_K)$ and $z = (b_{K0} - b_K, \ldots, b_{KJ} - b_K)$.

We next show $V^{(ANCOVA)} \succeq V^{(IMMRM)}$. By the definition of $\Sigma_j^{(ANCOVA)}$ and $\Sigma_j^{(IMMRM)}$, we have

$$\Sigma_j^{(ANCOVA)} - \Sigma_j^{(IMMRM)} = \{Cov(Y(j), X) - Cov*(Y, X)\}Var(X)^{-1}\{Cov(X, Y(j)) - Cov*(X, Y)\}$$

is positive semi-definite, and

$$e_K^\top\Sigma_j^{(ANCOVA)}e_K = e_K^\top\Sigma_j^{(IMMRM)}e_K + (b_{Kj} - b_K)^\top Var(X)(b_{Kj} - b_K).$$

Using Equations (11) and (14) and the fact that $Var(X) = E[Var(X|S)] + Var(E[X|S])$, we have

$$V^{(ANCOVA)} - \tilde{V}^{(IMMRM)}$$

$$= V^{(ANCOVA)} - \tilde{V}^{(ANCOVA)} + \tilde{V}^{(ANCOVA)} - \tilde{V}^{(IMMRM)}$$

$$\succeq V^{(ANCOVA)} - \tilde{V}^{(ANCOVA)} + \tilde{V}^{(ANCOVA)} - \tilde{V}^{(IMMRM)}$$

$$- L \left( diag\{P^*(M_K = 1)^{-1}\pi_j^{-1}e_K^\top\Sigma_j^{(IMMRM)}e_K : j = 0, \ldots, J\} + r^\top Var(X)r \right) L^\top$$

$$= V^{(ANCOVA)} - \tilde{V}^{(ANCOVA)}$$

$$+ \frac{1}{P^*(M_K = 1)}L[diag\{\pi_j^{-1}(b_{Kj} - b_K)^\top Var(X)(b_{Kj} - b_K) : j = 0, \ldots, J\}]L^\top - Lr^\top Var(X)rL^\top$$

$$\succeq L[diag\{\pi_j^{-1}(b_{Kj} - b_K)^\top E[Var(X|S)](b_{Kj} - b_K) : j = 0, \ldots, J\}] - z^\top E[Var(X|S)]zL^\top$$

$$= LU^\top \{ (diag\{\pi\} - \pi\pi^\top) \otimes I_p \} U L^\top,$$

(18)
where \( \pi = (\pi_0, \ldots, \pi_J) \top \) and

\[
U^\top = \begin{pmatrix}
\pi_0^{-1}(b_{K0} - b_K)^\top E[\text{Var}(X|S)]^{\frac{1}{2}} \\
\vdots \\
\pi_J^{-1}(b_{KJ} - b_K)^\top E[\text{Var}(X|S)]^{\frac{1}{2}}
\end{pmatrix}.
\]

In the above derivation, the first \( \succeq \) results from Lemma 1 (2), the second \( \succeq \) comes from \( P^*(M_K = 1) \leq 1 \) and \( Lz^\top = Lr^\top \). Since \( \text{diag} \{ \pi \} \succeq \pi \pi^\top \), then \( (\text{diag} \{ \pi \} - \pi \pi^\top) \otimes I_p \) is positive semi-definite (by Lemma 2) and hence \( V^{(ANCOVA)} \succeq \tilde{V}^{(IMMRM)} \).

We next show \( V^{(MMRM-I)} \succeq V^{(IMMRM)} \). Using the definition of \( \Sigma_j^{(MMRM-I)} \) and \( \Sigma_j^{(IMMRM)} \), we have

\[
\Sigma_j^{(MMRM-I)} - \Sigma_j^{(IMMRM)} = \Lambda_j, \quad \Lambda_j = \left\{ \text{Cov}(Y(j), X) - \beta_x^\top \text{Var}(X) \right\} \text{Var}(X)^{-1} \left\{ \text{Cov}(X, Y(j)) - \text{Var}(X) \beta_x 1_K^\top \right\}
\]

is positive semi-definite. By Equations (12) and (14), we have

\[
\tilde{V}^{(MMRM-I)} - \tilde{V}^{(IMMRM)} = L \text{diag} \left\{ e_K^\top E^*[\Sigma_j^{(IMMRM)} + \Lambda_j] E^*[\Sigma_j]^{-1} e_K : j = 0, \ldots, J \right\} L^\top
\]

\[
\succeq L \text{diag} \left\{ e_K^\top E^*[\Sigma_j]^{-1} E^* \left[ \pi_j^{-1} \Sigma_j \Lambda_j \Sigma_j \right] E^*[\Sigma_j]^{-1} e_K : j = 0, \ldots, J \right\} L^\top - Lr^\top \text{Var}(X)^{-1} r L^\top
\]

\[
\succeq L \text{diag} \left\{ e_K^\top \left[ \beta_x - \beta_x \right]^\top \text{Var}(X) (\beta_x - \beta_x) : j = 0, \ldots, J \right\} L^\top - L v^\top \text{Var}(X)^{-1} v L^\top,
\]

where the first \( \succeq \) results from Lemma 1 (3), the second \( \succeq \) results from Lemma 1 (4) and the last equation comes from \( e_K^\top \Lambda_j e_K = (b_{Kj} - \beta_x)^\top \text{Var}(X) (b_{Kj} - \beta_x) \) and
\( \mathbf{L} \mathbf{v}^\top = \mathbf{L} \mathbf{r}^\top \). By Equation (15) and \( \text{Var}(\mathbf{X}) = E[\text{Var}(\mathbf{X}|S)] + \text{Var}(E[\mathbf{X}|S]) \), we have

\[
\mathbf{V}^{(\text{MMRM-I})} - \tilde{\mathbf{V}}^{(\text{IMMRM})} = \mathbf{V}^{(\text{MMRM-I})} - \tilde{\mathbf{V}}^{(\text{MMRM-I})} + \tilde{\mathbf{V}}^{(\text{MMRM-I})} - \tilde{\mathbf{V}}^{(\text{IMMRM})} \\
\succeq - \mathbf{L}[\text{diag}\{\pi_j^{-1}(\mathbf{b}_{kj} - \underline{\beta}_k)^\top \text{Var}(\mathbf{X}|S))(\mathbf{b}_{kj} - \underline{\beta}_k) : j = 0, \ldots, J]\} + v \text{Var}(E[\mathbf{X}|S]) v^\top | \mathbf{L}^\top \\
+ \mathbf{L}[\text{diag}\{\pi_j^{-1}(\mathbf{b}_{kj} - \underline{\beta}_k)^\top \text{Var}(\mathbf{X})(\mathbf{b}_{kj} - \underline{\beta}_k) : j = 0, \ldots, J\}] \mathbf{L}^\top - \mathbf{L} v^\top \text{Var}(\mathbf{X}) v \mathbf{L}^\top \\
= \mathbf{L} [\text{diag}\{\pi_j^{-1}(\mathbf{b}_{kj} - \underline{\beta}_k)^\top \text{Var}(\mathbf{X}|S))(\mathbf{b}_{kj} - \underline{\beta}_k) : j = 0, \ldots, J\}] - v E[\text{Var}(\mathbf{X}|S)] v^\top | \mathbf{L}^\top \\
= \mathbf{L} Z^\top \{(\text{diag}\{\pi\} - \pi \pi^\top) \otimes \mathbf{I}_p\} \mathbf{Z}^\top, \\
(19)
\]

where

\[
Z^\top = \begin{pmatrix} 
\pi_0^{-1}(\mathbf{b}_{k0} - \underline{\beta}_k)^\top E[\text{Var}(\mathbf{X}|S)] \frac{1}{2} \\
\vdots \\
\pi_j^{-1}(\mathbf{b}_{kj} - \underline{\beta}_k)^\top E[\text{Var}(\mathbf{X}|S)] \frac{1}{2}
\end{pmatrix}.
\]

Since \( \text{diag}\{\pi\} \succeq \pi \pi^\top \), then we get \( \mathbf{V}^{(\text{MMRM-I})} \succeq \tilde{\mathbf{V}}^{(\text{IMMRM})} = \mathbf{V}^{(\text{IMMRM})} \).

Next, for showing \( \mathbf{V}^{(\text{MMRM-II})} \succeq \mathbf{V}^{(\text{IMMRM})} \), we can follow a similar proof as in the previous paragraph, where \( \Sigma^{(\text{MMRM-I})} \) is substituted by \( \Sigma^{(\text{MMRM-II})} \), and get

\[
\mathbf{V}^{(\text{MMRM-II})} - \tilde{\mathbf{V}}^{(\text{IMMRM})} \succeq \mathbf{L} \mathbf{U}^\top \{(\text{diag}\{\pi\} - \pi \pi^\top) \otimes \mathbf{I}_p\} \mathbf{U} \mathbf{L}^\top,
\]

which is positive semi-definite.

Finally, we give the necessary and sufficient conditions for \( \mathbf{V}^{(\text{est})} = \mathbf{V}^{(\text{IMMRM})} \), \((\text{est}) \in \{\text{ANCOVA}, \text{MMRM-I}, \text{MMRM-II}\}\) in Proposition 1 below.

**Proposition 1.** Assume \( K > 1 \), Assumption 1 and regularity conditions in the Supplementary Material. For \( t = 1, \ldots, K \) and \( j = 0, \ldots, J \), we denote \( \mathbf{b}_{tj} = \text{Var}(\mathbf{X})^{-1} \text{Cov}\{\mathbf{X}, Y_t(j)\} \).

Then \( \mathbf{V}^{(\text{ANCOVA})} = \mathbf{V}^{(\text{IMMRM})} \) if and only if either of the following two sets of conditions (a-b) holds:
(a) for each $j = 0, \ldots, J$, $P(M_K(j) = 1) = 1$ and 
\[(1 - 2I\{J = 1\}\pi_0)(b_{Kj} - b_{K0})^\top E[\text{Var}(X|S)] = 0;\]

(b) for each $t = 1, \ldots, K - 1$ and $j = 0, \ldots, J$, 
\[P(M_t(j) = 1, M_K(j) = 0) \text{ Cov}\{Y_t(j) - b_{Kj}^\top X, Y_K(j) - b_{Kj}^\top X\} = 0 \text{ and } b_{Kj} = b_{K0}.\]

In addition, $V^{(MMRM-I)} = V^{(IMMRM)}$ if and only if

(a') for $j = 0, \ldots, J$ and $m \in \{0, 1\}^K \setminus \{0_K\}$ with $P^*(M = m)$, $e_K^\top \{E[V_{jm}] - E[^*V_{jm}]\} = 0,$

(b') for $j = 0, \ldots, J$ and $m \in \{0, 1\}^K \setminus \{0_K\}$ with $P^*(M = m)$, $e_K^\top E[V_{M}]^{-1}V_{m}A_j$ is a constant vector,

(c') for $j = 0, \ldots, J$, $[b_{Kj} - \beta_{(MMRM-I)}^X - I\{J = 1\}\pi_j(b_{Kj} - b_{K0})]^\top E[\text{Var}(X|S)] = 0,$

where $\beta_{(MMRM-I)}^X$ is the probability limit of $\beta^X$ in the MMRM-I working model.

Also, $V^{(MMRM-II)} = V^{(IMMRM)}$ if and only if

(a'') for $j = 0, \ldots, J$ and $m \in \{0, 1\}^K \setminus \{0_K\}$ with $P^*(M = m)$, $e_K^\top \{E[V_{jm}] - E[^*V_{jm}]\} = 0,$

(b'') for $j = 0, \ldots, J$ and $m \in \{0, 1\}^K \setminus \{0_K\}$ with $P^*(M = m)$, 
\[e_K^\top E[V_{M}]^{-1}V_{m}\{\text{Cov}(Y(j), X) - \text{Cov}(Y, X)\}V_{M}^{-1}\{\text{Cov}(X, Y(j)) - \text{Cov}(X, Y)\} \]

is a constant vector,

(c'') for $j = 0, \ldots, J$, $(1 - 2I\{J = 1\}\pi_0)(b_{Kj} - b_{K0})^\top E[\text{Var}(X|S)] = 0.$

Proof. We first derive the necessary and sufficient conditions for $V^{(ANCOVA)} = V^{(IMMRM)}$. Recall the derivation, i.e., Equations (18), for showing $V^{(ANCOVA)} \succeq V^{(IMMRM)}$ in the proof.
of Theorem 1. By check the two inequalities and the last row in Equations (18). We have 
\[ V^{(ANCOVA)} = V^{(IMMRM)} \]  if and only if the following three conditions hold:

(i) \[ L \text{ diag}\{ \pi_j^{-1} e_K^T (\pi_j^{-1} \sum_j^{(IMMRM)} e_K - E^*[V_j M])^{-1} e_K : j = 0, \ldots, J \} L = 0, \]

(ii) \[ \{1 - P^*(M_K = 1)\} L \text{ diag}\{ \pi_j^{-1} (b_{Kj} - b_K)^T \text{Var}(X) (b_{Kj} - b_K) : j = 0, \ldots, J \} L^T = 0, \]

(iii) \[ LU^T\{(\text{diag}\{\pi\} - \pi\pi^T) \otimes I_p\} UL^T = 0. \]

For Condition (i), Lemma 1 (2) and the assumption that \( K > 1 \) imply that the equation holds if and only if \( P^*(M_K = 0, M_t = 1) e_t^T \sum_j^{(IMMRM)} e_K = 0 \) for \( t = 1, \ldots, K - 1 \) and \( j = 0, \ldots, J \). Equation (10) implies that \( e_t^T \sum_j^{(IMMRM)} e_K = \text{Cov}(Y_t(j) - b_{tj}^T X, Y_K(j) - b_{Kj}^T X). \) The MCAR assumption implies that \( P^*(M_K = 0, M_t = 1) = P(M_K(j) = 0, M_t(j) = 1) \) for \( j = 0, \ldots, J \). Hence Condition (i) is equivalent to

(i) \[ P(M_K(j) = 0, M_t(j) = 1) \text{Cov}(Y_t(j) - b_{tj}^T X, Y_K(j) - b_{Kj}^T X) \] for \( t = 1, \ldots, K - 1 \) and \( j = 0, \ldots, J \).

Condition (ii) is equivalent to

(ii) \[ P^*(M_K = 1) \text{ or } \text{Cov}(Y_K(j) - Y_K(0), X) = 0. \]

For Condition (iii), since \( LU^T\{(\text{diag}\{\pi\} - \pi\pi^T) \otimes I_p\} UL^T \) is positive semi-definite, then it is \( 0 \) if and only if all of its diagonal entries are 0. Denoting \( u_j = E[\text{Var}(X|S)]^{1/2} (b_{Kj} - b_K) \), we get that the \((j, j)\)-th entry of matrix \( LU_{S=s}(\text{diag}\{\pi\} - \pi\pi^T) U_{S=s} L^T \) is

\[
\pi_j^{-1} u_j^T u_j + \pi_0^{-1} u_0^T u_0 - (u_j - u_0)^T (u_j - u_0) \]

\[ = \frac{1}{\pi_0 \pi_j} (\pi_0 u_j + \pi_j u_0)^T (\pi_0 u_j + \pi_j u_0) + (1 - \pi_0 - \pi_j) \left( \frac{1}{\pi_j} u_j^T u_j + \frac{1}{\pi_0} u_0^T u_0 \right), \]

28
which is equal to 0 if and only if either \( u_0 = u_j = 0 \), or \( \pi_0 + \pi_j = 1 \) and \( \pi_0 u_j + \pi_j u_0 = 0 \). The former case is equivalent to \( E[Var(X|S)] \frac{1}{2} (b_{Kj} - b_{K0}) = 0 \) for \( j = 1, \ldots, K \); and the later case is equivalent to \( J = 1 \) and \( (\pi_1 - \pi_0)E[Var(X|S)] \frac{1}{2} (b_{K1} - b_K) \). Hence Condition (iii) is equivalent to

\[
(iii) \quad (I\{J = 1\}(\pi_j - \pi_0) + I\{J > 1\})E[Var(X|S)](b_{Kj} - b_K) = 0 \quad \text{for} \quad j = 1, \ldots, J.
\]

Combining Conditions (i-iii) together, we observe that, \( P^*(M_K = 1) = 1 \) in Condition (ii) implies Condition (i), and \( Cov(Y_K(j) - Y_K(0), X) = 0 \) in Condition (ii) implies Condition (iii). As a result, the three conditions can be summarized into two conditions, which are

(a) for each \( j = 1, \ldots, J \), \( P(M_K(j) = 1) = 1 \) and

\[
(1 - 2I\{J = 1\})\pi_0(b_{Kj} - b_{K0})^\top E[Var(X|S)] = 0;
\]

(b) for each \( t = 1, \ldots, K - 1 \) and \( j = 0, \ldots, J \),

\[
P(M_t(j) = 1, M_K(j) = 0) \quad Cov\{Y_t(j) - b_{tj}^\top X, Y_K(j) - b_{Kj}^\top X\} = 0 \quad \text{and} \quad b_{Kj} = b_{K0}.
\]

For the MMRM-I estimator, the derivations, i.e., Equations (19), imply that \( V^{(MMRM-I)} = V^{(MMRMR)} \) if and only if the following conditions hold:

(i') for \( j = 0, \ldots, J \) and \( m \in \{0,1\}^K \setminus \{0_K\} \) with \( P^*(M = m) \), \( e_K^\top \{E[V_{jm}]^{-1}V_{jm} - E^*[V_M]^{-1}V_m\} = 0 \),

(ii') for \( j = 0, \ldots, J \) and \( m \in \{0,1\}^K \setminus \{0_K\} \) with \( P^*(M = m) \), \( e_K^\top E[V_M]^{-1}V_m A_j \) is a constant vector,

(iii') \( LZ^\top \{(diag(\pi) - \pi \pi^\top) \otimes I_p\} ZL^\top = 0 \).

Similar to the analysis for Condition (iii), Condition (iii') is equivalent to
(iii') for $j = 1, \ldots, J$, $[b_{kj} - \theta^{(\text{MMRM-I})}_X] - I\{J = 1\} \pi_j (b_{kj} - b_{k0})^\top E[Var(X|S)] = 0$, which is the necessary condition given in Corollary 1.

For the MMRM-II estimator, similarly, we have $V^{(\text{MMRM-I})} = V^{(\text{IMMRM})}$ if and only if the following conditions hold:

(i”) for $j = 0, \ldots, J$ and $m \in \{0, 1\} \setminus \{0_K\}$ with $P^*(M = m)$, $e^\top_K \{E[\Sigma_{jm}]^{-1}\Sigma_{jm} - E^*\Sigma_{Mm}^{-1}\Sigma_{Mm}\} = 0$,

(ii”) for $j = 0, \ldots, J$ and $m \in \{0, 1\} \setminus \{0_K\}$ with $P^*(M = m)$,

$e^\top_K E[\Sigma_{Mm}]^{-1}\Sigma_{Mm} \{Cov(Y(j), X) - Cov^*(Y, X)\} Var(X)^{-1} \{Cov(X, Y(j)) - Cov^*(X, Y)\}$

is a constant vector,

(iii”) $LU^\top \{(\text{diag}\{\pi\} - \pi\pi^\top) \otimes I_p\} UL^\top = 0$.

The Condition (iii”) is the same as Condition (iii), which is the necessary condition shown in Corollary 1.

\[\Box\]

Lemma 5. Assume the same assumption as in Theorem 1. Then the influence functions of $\hat{\Delta}^{(\text{ANCOVA})}$, $\hat{\Delta}^{(\text{MMRM-I})}$ and $\hat{\Delta}^{(\text{MMRM-II})}$ are given by Equations (7), (8) and (9), respectively. Under simple randomization, the asymptotic covariance matrix of $\hat{\Delta}^{(\text{ANCOVA})}$, $\hat{\Delta}^{(\text{MMRM-I})}$ and $\hat{\Delta}^{(\text{MMRM-II})}$ are given by Equations (11), (12) and (13), respectively.

\[\text{Proof.}\] We first derive the influence function for the MMRM-I estimator. Theorem 1 of Wang et al. (2021) implies that $IF^{(\text{MMRM-I})}(A, X, Y, M; \theta) = B^{-1} \psi^{(\text{MMRM})}(A, X, Y, M; \theta)$, where $B = E^* \left[\frac{\partial}{\partial \theta} \psi^{(\text{MMRM})}(A, X, Y, M; \theta)\right]_{\theta = \theta}$. Using the formula (2) of $\psi^{(\text{MMRM})}$, we
can show that

\[
B = \begin{bmatrix}
-E^*[V_M] & -E^*[V_M (I_K \otimes A)^\top] & -E^*[V_M u(X)^\top] & 0 \\
-E^*[(I_K \otimes A)V_M] & -E^*[(I_K \otimes A)V_M (I_K \otimes A)^\top] & -E^*[(I_K \otimes A)V_M u(X)^\top] & 0 \\
-E^*[u(X)V_M] & -E^*[u(X)V_M (I_K \otimes A)^\top] & -E^*[u(X)V_M u(X)^\top] & B_{34} \\
0 & 0 & B_{34}^\top & B_{44}
\end{bmatrix},
\]

where \( B_{34} \in \mathbb{R}^{q \times r} \) and \( B_{44} \in \mathbb{R}^{r \times r} \) are matrices not related to the influence function of \( \hat{\Delta}^{(MMRM-I)} \). The zeros in the above matrix result from the following derivation:

\[
E^* \left[ V_M \frac{\partial \Sigma}{\partial \alpha_j} V_M (Y - Q^\top \beta) \right] = E^* \left[ V_M \frac{\partial \Sigma}{\partial \alpha_j} V_M \right] E^* [Y - \beta_0 - (I_K \otimes A)^\top \beta_A - u(X)^\top \beta_{u(X)}] = 0,
\]

and, similarly, \( E^* \left[ (I_K \otimes A)V_M \frac{\partial \Sigma}{\partial \alpha_j} V_M (Y - Q^\top \beta) \right] = 0 \). By the regularity conditions, \( B \) is invertible. To compute \( B^{-1} \), we define

\[
D = \begin{pmatrix}
I_K & 0 & 0 & 0 \\
-E^*[I_K \otimes A] & I_K & 0 & 0 \\
-E^*[u(X)] & 0 & I_q & 0 \\
0 & 0 & 0 & I_r
\end{pmatrix},
\]

and \( F = DBD^\top \). Since \( D \) is a lower triangular matrix and hence invertible, then \( F \) is invertible. Since MCAR implies that \( E^*[u(X)V_M] = E^*[u(X)]E^*[V_M] \) and \( E^*[(I_K \otimes A)V_M] = E^*[I_K \otimes A]E^*[V_M] \), then

\[
F = \begin{pmatrix}
-E^*[V_M] & 0 & 0 & 0 \\
0 & -E^*[V_M] \otimes Var^*(A) & 0 & 0 \\
0 & 0 & -Var^*(u(X))E^*[V_M]^2 & B_{34} \\
0 & 0 & B_{34}^\top & B_{44}
\end{pmatrix},
\]
where \( \text{Var}^*(A) = E^*[AA^\top] - E^*[A]E^*[A]^\top \), \( F_{33} \in \mathbb{R}^{q \times q} \) is a matrix not related to the influence function of \( \tilde{\beta}_A \). Then

\[
B^{-1} = D^\top F^{-1} D
\]

\[
= D^\top \begin{bmatrix}
-E^*[V_M]^{-1} & 0 & 0 & 0 \\
0 & -E^*[V_M]^{-1} \otimes \text{Var}^*(A)^{-1} & 0 & 0 \\
0 & 0 & \tilde{B}_{33} & \tilde{B}_{34} \\
0 & 0 & \tilde{B}_{34}^\top & \tilde{B}_{44}
\end{bmatrix} D
\]

\[
= \begin{bmatrix}
\tilde{B}_{11} & C^\top & -E^*[u(X)^\top]\tilde{B}_{33} & -E^*[u(X)^\top]\tilde{B}_{34} \\
C & -E^*[V_M]^{-1} \otimes \text{Var}^*(A)^{-1} & 0 & 0 \\
-\tilde{B}_{33}E^*[u(X)] & 0 & \tilde{B}_{33} & \tilde{B}_{34} \\
-\tilde{B}_{34}^\top E^*[u(X)] & 0 & \tilde{B}_{34}^\top & \tilde{B}_{44}
\end{bmatrix}
\]

where \( C = \{E^*[V_M]^{-1} \otimes \text{Var}^*(A)^{-1}\} E^*[I_K \otimes A] \) and \( \tilde{B}_{11} \in \mathbb{R}^{K \times K}, \tilde{B}_{33} \in \mathbb{R}^{q \times q}, \tilde{B}_{34} \in \mathbb{R}^{q \times r} \) and \( \tilde{B}_{44} \in \mathbb{R}^{r \times r} \) are matrices that are not related to the influence function of \( \tilde{\beta}_A \) (as shown below). Since \( \beta_A \) are the \((K+1)\)-th, \ldots, \((J+1)K\)-th entries in \( \theta \), we need the \((K+1)\)-th, \ldots, \((J+1)K\)-th rows of \( B^{-1} \) to derive the influence function for \( \tilde{\beta}_A \), which are

\[
[C - \{E^*[V_M]^{-1} \otimes \text{Var}^*(A)^{-1}\} 0 0]
\]

Then the influence function for \( \tilde{\beta}_A \) is

\[
\{E^*[V_M]^{-1} \otimes \text{Var}^*(A)^{-1}\} \{I_K \otimes (A - E^*[A])\} V_M (Y - Q^\top \beta),
\]

which implies that the influence function for \( \Delta^{(MMRM-I)} \) is

\[
IF^{(MMRM-I)} = \text{Var}^*(A)^{-1} (A - E^*[A]) e_K^\top E^*[V_M]^{-1} V_M (Y - Q^\top \beta).
\]
Since $Var^*(A)^{-1}(A - E^*[A]) = L(I(\{A=0\}/\pi_0), \ldots, I(\{A=I\}/\pi_J))^\top$, we get the desired formula of $IF^{(MMRM-I)}$.

We next compute $Y - Q^\top \beta$. By $E^*[\psi^{(MMRM)}(A, X, Y, M; \theta)] = 0$, we have $\beta = E^*[Q \Sigma M Q]^{-1} E^*[Q \Sigma M Y]$. Recalling $u(X) = X 1_K^\top$ for the MMRM-I estimator and following a similar procedure for calculating $B^{-1}$, we have

$$
\beta = \left( \begin{array}{ccc}
I_K & 0 & 0 \\
-E^*[I_K \otimes A] & I_K & 0 \\
-E^*[u(X)] & 0 & I_q \\
\end{array} \right)^\top \left( \begin{array}{c}
E^*[\Sigma M]^{-1} \\
0 \\
0 \\
\end{array} \right)
\left( \begin{array}{c}
E^*[\Sigma M]^{-1} \otimes Var^*(A)^{-1} \\
0 \\
0 \\
\end{array} \right)
\left( \begin{array}{c}
\Sigma M \otimes \Sigma M \\
\Sigma M \otimes \Sigma M \\
\Sigma M \otimes \Sigma M \\
\end{array} \right)
\left( \begin{array}{c}
E^*[u(X) E^*[\Sigma M]^{-1/2}] \\
0 \\
0 \\
\end{array} \right)^{-1}
$$

which implies $\beta_A = E^*[\tilde{Y} \otimes \tilde{A}]$ and $\beta_{\Sigma X} = \frac{1}{k E^*[\Sigma M] I_K} Cov^*(Y, X) Var(X)^{-1}$. Since $\beta_0$ satisfies $E^*[Y - \beta_0 - (I_K \otimes A)^\top \beta_A - u(X)^\top \beta_{u(X)}] = 0$, we get

$$
Y - Q^\top \beta = \tilde{Y} - \beta_A^\top (I_K \otimes \tilde{A}) - 1_K \beta_A^\top X \tilde{X}.
$$

Then direct calculation gives the desired formula of $Y - Q^\top \beta$.

We next calculate $\tilde{\Sigma}^{(MMRM-I)}$. Since $\Sigma$ is unstructured, the second set of estimating equations $\psi^{(MMRM)}$ implies that, for each $r, s = 1, \ldots, K$ and $j = 0, \ldots, J$, we have

$$
0 = E^* [-\text{tr}(\Sigma M (e_r e_s^\top + e_s e_r^\top)) + (Y - Q^\top \beta)^\top \Sigma M (e_r e_s^\top + e_s e_r^\top)]\Sigma M (Y - Q^\top \beta)
= E^* [2e_r^\top \{ -\Sigma M + \Sigma M (Y - Q^\top \beta)(Y - Q^\top \beta)^\top \Sigma M \} e_s],
$$

which implies that, for $j = 0, \ldots, J$, $E^*[\Sigma M] + E^*[\Sigma M (Y - Q^\top \beta)(Y - Q^\top \beta)^\top \Sigma M] = 0$.

$$
E^*[\Sigma M] + E^*[\Sigma M (Y - Q^\top \beta)(Y - Q^\top \beta)^\top \Sigma M] = 0.
$$
Since $E^*[\mathbf{V}_M] = E^*[\mathbf{V}_M \Sigma^{(MMRM-I)} \mathbf{V}_M]$ and the regularity condition (2) implies that $\Sigma^{(MMRM-I)} = E^*[(\mathbf{Y} - Q^\top \beta)(\mathbf{Y} - Q^\top \beta)^\top]$. Thus,

\[
\tilde{\mathbf{V}}^{(MMRM-I)} = E[I^F(\mathbf{MMRM-I}) I^F(\mathbf{MMRM-I})^\top] = L \, \text{diag}\left\{ e_k^\top E^*[\mathbf{V}_M]^{-1} E^* \left[ \frac{I\{A = j\}}{\pi_j^2} \mathbf{V}_M (\mathbf{Y} - Q^\top \beta)(\mathbf{Y} - Q^\top \beta)^\top \mathbf{V}_M \right] E^*[\mathbf{V}_M]^{-1} e_k : j = 0, \ldots, J \right\} L^\top.
\]

For the MMRM-II estimator, we can follow a similar proof for the MMRM-I estimator and get the desired influence function and asymptotic covariance matrix. The only difference comes from $u(\mathbf{X}) = \mathbf{I}_K \otimes \mathbf{X}$.

For the ANCOVA estimator, we observed that it is a special case of the MMRM-I estimator setting $K = 1$. Then we have $\mathbf{V}_M = I\{M_K = 1\}(e_k^\top \Sigma^{(ANCOVA)} e_k)^{-1}$, which naturally implies the desired influence function and asymptotic covariance matrix.

**Lemma 6.** Assume the same assumption as in Theorem 1 and assume $\Sigma_{j,j} = 0$ are unstructured. Then the influence function of $\hat{\Delta}^{(MMRM)}$ is given by Equation (10) and the asymptotic covariance matrix of $\hat{\Delta}^{(MMRM)}$ is given by Equation (14).

**Proof.** Following a similar procedure as in Lemma 5, we get that the influence function for $\hat{\Delta}^{(MMRM)}$ is $I^F(\mathbf{MMRM}) = U_1 + U_2$, where

\[
U_1 = (e_k^\top \mathbf{I}_J) H^{-1}(\mathbf{I}_K \otimes \mathbf{A} - E^*[\mathbf{V}_{AM} \otimes \mathbf{A}] E^*[\mathbf{V}_{AM}]^{-1}) \mathbf{V}_{AM}(\mathbf{Y} - R^\top \gamma)
\]
\[
U_2 = (e_k \otimes \mathbf{I}_J)^\top \beta_A + (e_k \otimes \mathbf{X} \otimes \mathbf{I}_J)^\top \beta_{AX} - \Delta^*.
\]
where \( \mathbf{H} = E^*[\mathbf{V}_{AM} \otimes \mathbf{A}^\top] - E^*[\mathbf{V}_{AM} \otimes \mathbf{A}]E^*[\mathbf{V}_{AM}]^{-1}E^*[\mathbf{V}_{AM} \otimes \mathbf{A}^\top] \).

We next compute \( \mathbf{U}_1 \). Define \( \mathbf{\delta}_j \in \mathbb{R}^J \) has the \( J \)-th entry 1 and the rest 0. We have

\[
E^*[\mathbf{V}_{AM} \otimes \mathbf{A}^\top]^{-1} = \left( \sum_{j=1}^J E[\pi_j \mathbf{V}_{jM}] \otimes \mathbf{\delta}_j \mathbf{\delta}_j^\top \right)^{-1} = \sum_{j=1}^J E[\pi_j \mathbf{V}_{jM}]^{-1} \otimes \mathbf{\delta}_j \mathbf{\delta}_j^\top
\]

and hence

\[
-E^*[\mathbf{V}_{AM}] + E^*[\mathbf{V}_{AM} \otimes \mathbf{A}^\top]E^*[\mathbf{V}_{AM} \otimes \mathbf{A}^\top]^{-1}E^*[\mathbf{V}_{AM} \otimes \mathbf{A}]
= -E[\mathbf{V}_{AM}] + \sum_{j=1}^J E[\pi_j \mathbf{V}_{jM}] \otimes \mathbf{\delta}_j^\top \mathbf{\delta}_j
= -E[\pi_0 \mathbf{V}_{0M}].
\]

Using the Woodbury matrix identity, we get

\[
\mathbf{H}^{-1} = E^*[\mathbf{V}_{AM} \otimes \mathbf{A}^\top]^{-1} - E^*[\mathbf{V}_{AM} \otimes \mathbf{A}^\top]^{-1}E^*[\mathbf{V}_{AM} \otimes \mathbf{A}]
\]

\[
(E^*[\mathbf{V}_{AM}] + E^*[\mathbf{V}_{AM} \otimes \mathbf{A}^\top]E^*[\mathbf{V}_{AM} \otimes \mathbf{A}^\top]^{-1}E^*[\mathbf{V}_{AM} \otimes \mathbf{A}])^{-1}
\]

\[
E^*[\mathbf{V}_{AM} \otimes \mathbf{A}^\top]E^*[\mathbf{V}_{AM} \otimes \mathbf{A}^\top]^{-1}
\]

\[
= \sum_{j=1}^J E[\pi_j \mathbf{V}_{jM}]^{-1} \otimes \mathbf{\delta}_j \mathbf{\delta}_j^\top - (\mathbf{I}_K \otimes \mathbf{1}_J)(-E[\pi_0 \mathbf{V}_{0M}])^{-1}(\mathbf{I}_K \otimes \mathbf{1}_J)
\]

\[
= \sum_{j=1}^J E[\pi_j \mathbf{V}_{jM}]^{-1} \otimes \mathbf{\delta}_j \mathbf{\delta}_j^\top + E[\pi_0 \mathbf{V}_{0M}]^{-1} \otimes \mathbf{1}_J \mathbf{1}_J^\top
\]
Then we get
\[
H^{-1}(I_K \otimes A - E^* [V_{AM} \otimes A] E^* [V_{AM}]^{-1})
= \left( \sum_{j=1}^{J} E[\pi_{j} V_{jM}]^{-1} \otimes \delta_j \delta_j^T + E[\pi_0 V_{0M}]^{-1} \otimes 1_j 1_j^T \right)
\left( I_K \otimes A - \sum_{j=1}^{J} E[\pi_j V_{jM}] E^* [V_{AM}]^{-1} \otimes \delta_j \right)
= \sum_{j=1}^{J} E[\pi_{j} V_{jM}]^{-1} \otimes \delta_j I\{A = j\} + E[\pi_0 V_{0M}]^{-1} \otimes 1_j(1 - I\{A = 0\}) - \sum_{j=1}^{J} E^* [V_{AM}]^{-1} \otimes 1_j
- E[\pi_0 V_{0M}]^{-1}(E^* [V_{AM}] - E[\pi_0 V_{0M}]) E^* [V_{AM}]^{-1} \otimes 1_j
= \sum_{j=1}^{J} E[\pi_{j} V_{jM}]^{-1} \otimes \delta_j I\{A = j\} - E[\pi_0 V_{0M}]^{-1} \otimes 1_j I\{A = 0\},
\]
which implies that \( U_1 = LT^{(IMMRM)} V_{AM} (Y - R^T \gamma) \). By \( E^*[\psi^{(IMMRM)}(A, X, Y, M; \theta)] = 0 \), we have \( \gamma = E^* [R V_{AM} R]^{-1} E^* [R V_{AM} Y] \). Following a similar procedure for calculating \( \beta \) in Lemma 5, we get
\[
Y - R^T \gamma = \sum_{j=1}^{J} I\{A = j\} \{ Y(j) - E[Y(j)] - Cov(Y(j), X) Var(X)^{-1} (X - E[X]) \}.
\]

We next compute \( U_2 \). For each \( j = 1, \ldots, J \), we have the \( j \)-th entry of \( U_2 \) is
\[
\delta_j^T U_2 = \beta_{A_{jK}} + \beta_{A_{X_{jK}}}^T X - \Delta_j^*.
\]
Since we have shown the model-robustness of \( \Delta^{(IMMRM)} \) in the proof of Theorem 1, then \( \Delta_j^* = \beta_{A_{jK}} + \beta_{A_{X_{jK}}}^T E[X] \) and hence \( \delta_j^T U_2 = \beta_{A_{X_{jK}}}^T (X - E[X]) \). By the formula of \( Y - R^T \gamma \), we have
\[
\beta_{A_{X_{jK}}}^T (X - E[X]) = (I\{A = j\} - I\{A = 0\}) e^T_K (Y - R^T \gamma)
= Cov(Y_K(j) - Y_K(0), X) Var(X)^{-1} (X - E[X])
= (b_{Kj} - b_{K0})^T (X - E[X]),
\]
36
which implies $U_2 = Lr^T(X - E[X])$.

We next compute $\tilde{V}^{(IMMRM)}$, which is

$$\tilde{V}^{(IMMRM)} = E^*[IF^{(IMMRM)}IF^{(IMMRM)^T}] = E^*[U_1U_1^T] + E^*[U_1U_2^T] + E^*[U_2U_1^T] + E^*[U_2U_2^T].$$

For $E^*[U_1U_1^T]$, since $E^*[\psi^{(IMMRM)}(A, X, Y, M)] = 0$ imply that $E^*[V_{AM}(Y - R^T \gamma)|X] = 0$ and $E^*[(1 \otimes A)V_{AM}(Y - R^T \gamma)|X] = 0$, then $E^*[U_1|X] = 0$ and hence $E^*[U_1U_1^T] = E^*[E[U_1|X]U_1^T] = 0$. Thus, $\tilde{V}^{(IMMRM)} = E^*[U_1U_1^T] + E^*[U_2U_2^T]$. For $E^*[U_1U_1^T]$, since $\Sigma_j$ is unstructured for each $j = 0, \ldots, J$, the second set of estimating equations $\psi^{(IMMRM)}$ implies that, for each $r, s = 1, \ldots, K$ and $j = 0, \ldots, J$, we have

$$0 = E^*[-\text{tr}(V_{AM}I\{A = j\}(e_r e_s^T + e_s e_r^T)) + (Y - R^T \gamma)^T V_{AM}I\{A = j\}(e_r e_s^T + e_s e_r^T)V_{AM}(Y - R^T \gamma)]
= E^*[2I\{A = j\}e_r^T(-V_{AM} + V_{AM}(Y - R^T \gamma)(Y - R^T \gamma)^T V_{AM})e_s],$$

which implies that, for $j = 0, \ldots, J$,

$$E[-\pi_j V_{jM}] + E^*[V_{jM}I\{A = j\}(Y - R^T \gamma)(Y - R^T \gamma)^T V_{jM}] = 0. \quad (20)$$

Hence

$$E^*[U_1U_1^T] = E^*[LT^{(IMMRM)}V_{AM}(Y - R^T \gamma)(Y - R^T \gamma)^T V_{AM}T^{(IMMRM)}L^T]
= L \text{ diag} \left\{ \frac{1}{\pi_j^2} e_k^T E[V_{jM}]^{-1} E^*[V_{AM}I\{A = j\}(Y - R^T \gamma)
(Y - R^T \gamma)^T V_{AM}]E[V_{jM}]^{-1} e_k : j = 0, \ldots, J \right\} L^T
= L \text{ diag} \{e_k^T E[\pi_j V_{jM}]^{-1} e_k : j = 0, \ldots, J \} L^T,$$

where the last equation results from Equation (20). Since $E^*[U_2U_2^T] = Lr^TVar(X)rL^T$, we get the desired formula of $\tilde{V}^{(IMMRM)}$.\qed
E.2 Proof of Corollary 1

Proof. The ANHECOVA estimator is a special case of the IMMRM model setting $K = 1$. Hence the consistency and asymptotic normality under simple or stratified randomization are implied by Theorem 1. Furthermore, by Equation (14), we have

$$\tilde{V}^{(ANHECOVA)} = \mathbf{L} \left( \text{diag}\{\pi_j^{-1}e^K e_j : j = 0, \ldots, J\} + \mathbf{r}^\top \text{Var}(\mathbf{X}) \mathbf{r} \right) \mathbf{L}.$$ 

Then, by Lemma 1 (2),

$$\tilde{V}^{(ANHECOVA)} - \tilde{V}^{(IMMRM)} = \mathbf{L} \text{diag}\{\pi_j^{-1}e^K (P^*(M_K = 1)^{-1} \Sigma^{(IMMRM)} - E[\mathbf{V}_M^{-1}]^{-1}) e_j : j = 0, \ldots, J\} \mathbf{L} \succeq 0,$$

with equality holds if and only if $P(M_t(j) = 1, M_K(j) = 0) \text{Cov}\{Y_t(j) - b_j^\top X, Y_K(j) - b_{Kj}^\top X\} = 0$ for each $t = 1, \ldots, K - 1$ and $j = 0, \ldots, J$. The final result comes from the fact that $\text{Cov}\{b_{ij}^\top X, Y_K(j) - b_{Kj}^\top X\} = 0$. $\square$

E.3 Proof of Corollary 2

Proof. By Equations (16) and (17), we have

$$\tilde{V}^{(ANCOVA)} - \tilde{V}^{(ANCOVA)} = \tilde{V}^{(MMRM-II)} - \tilde{V}^{(MMRM-II)} = \mathbf{L} \text{diag}\{\pi_j^{-1}(b_{Kj} - b_K)^\top \text{Var}(E[\mathbf{X}|S])(b_{Kj} - b_K) : j = 0, \ldots, J\} - \mathbf{z}^\top \text{Var}(E[\mathbf{X}|S]) \mathbf{z} \mathbf{L}^\top.$$ 

38
If $J = 1$ and $\pi_1 = \pi_0 = 0.5$, then $L = (-1, 1)$, $b_K = 0.5(b_{K1} + b_{K0})$ and $zL^\top = b_{K1} - b_{K0}$.

Hence

\[
\tilde{V}^{(ANCOVA)} - V^{(ANCOVA)} = \sum_{j=0}^{1} 2\{b_{K1} - 0.5(b_{K1} + b_{K0})\}^\top \text{Var}(E[X|S])\{b_{K1} - 0.5(b_{K1} + b_{K0})\} - (b_{K1} - b_{K0})^\top \text{Var}(E[X|S])(b_{K1} - b_{K0}) = 0.
\]

We next compare $V^{(ANCOVA)}$ and $V^{(MMRM-II)}$. By the definition of $\Sigma^{(ANCOVA)}$ and $\Sigma_j^{(ANCOVA)}$, we have $2\Sigma^{(ANCOVA)} = \Sigma_0^{(ANCOVA)} + \Sigma_1^{(ANCOVA)}$. Then Equation (11) implies that $\tilde{V}^{(ANCOVA)} = 4P^*(M_K = 1)^{-1}e_K^\top \Sigma^{(ANCOVA)} e_K$. Similarly, we have $\tilde{V}^{(MMRM-II)} = 4e_K^\top E^*[V_M]^{-1}e_K$. Since $\Sigma^{(ANCOVA)} = \Sigma^{(MMRM-II)}$, then Lemma 1 (2) implies that $\tilde{V}^{(ANCOVA)} - \tilde{V}^{(MMRM-II)} \geq e_K^\top (\Sigma^{(ANCOVA)} - \Sigma^{(MMRM-II)}) e_K = 0$.

Finally, we show $V^{(MMRM-I)} \geq V^{(MMRM-II)}$. Under two-armed equal randomization, Equation (15) implies that

\[
\tilde{V}^{(MMRM-I)} - V^{(MMRM-I)} = 4\{b_K - \beta_{X}\}^\top \text{Var}(E[X|S])\{b_K - \beta_{X}\}.
\]

In addition, since $2\Sigma^{(MMRM-I)} = \Sigma_0^{(MMRM-I)} + \Sigma_1^{(MMRM-I)}$, Equation (12) implies that $\tilde{V}^{(MMRM-I)} = 4e_K^\top E^*[V_M]^{-1}e_K$. By the definition of $\Sigma^{(MMRM-I)}$ and $\Sigma^{(MMRM-II)}$, we have

\[
e_K^\top (\Sigma^{(MMRM-I)} - \Sigma^{(MMRM-II)}) e_K = Var^*(Y_K) + \beta_{X}^\top \text{Var}(X)\beta_{X} - \beta_{X}^\top \text{Cov}^*(X, Y_K) - \text{Cov}^*(Y_K, X)\beta_{X} - Var^*(Y_K) - b_K^\top \text{Var}(X)b_K = \{b_K - \beta_{X}\}^\top \text{Var}(X)\{b_K - \beta_{X}\}.
\]
Hence

\[ V^{(MMRM-I)} - V^{(MMRM-II)} = V^{(MMRM-I)} - \tilde{V}^{(MMRM-I)} + \tilde{V}^{(MMRM-I)} - V^{(MMRM-II)} \]

\[ = -4\{b_K - \beta X\}^T Var(E[X|S])\{b_K - \beta X\} + 4e_K^T(E^*[\tilde{V}_M]^{-1} - E^*[\tilde{V}_M])e_K \]

\[ \geq -4\{b_K - \beta X\}^T Var(E[X|S])\{b_K - \beta X\} + 4e_K^T(\Sigma^{(MMRM-I)} - \Sigma^{(MMRM-II)})e_K \]

\[ = 4\{b_K - \beta X\}^T E[Var(X|S)]\{b_K - \beta X\} \]

\[ \geq 0, \]

where the inequality comes from Lemma 1 (5). \[ \square \]

F A counterexample showing MMRM-II is less precise than ANCOVA

We assume Assumption 1, simple randomization, \( \pi_1 = \frac{1}{3}, \pi_0 = \frac{2}{3} \), and

\[ E[Y(1)] = E[Y(0)] = 0, \]

\[ Var(Y(1)) = \begin{pmatrix} 4 & -3 \\ -3 & 4 \end{pmatrix}, \quad Var(Y(0)) = \begin{pmatrix} 4 & 3 \\ 3 & 4 \end{pmatrix}, \]

\[ p_{(1,0)} = p_{(0,1)} = p_{(1,1)} = \frac{1}{3}, \]

where \( p_m = P^*(M = m) \) for \( m \in \{0, 1\}^2 \). Then we have \( \Sigma_{i}^{(ANCOVA)} = \Sigma_{i}^{(MMRM-II)} = Var(Y(j)) \) for \( j = 0, 1 \). Furthermore,

\[ \Sigma^{(MMRM-II)} = \pi_0 \Sigma_0^{(MMRM-II)} + \pi_1 \Sigma_1^{(MMRM-II)} = \begin{pmatrix} 4 & 1 \\ 1 & 4 \end{pmatrix} \]

40
We define 
\[ C = \pi_0^{-1} \Sigma_0^{(ANCOVA)} + \pi_1^{-1} \Sigma_1^{(ANCOVA)} = \frac{9}{2} \begin{pmatrix} 4 & -1 \\ -1 & 4 \end{pmatrix}. \]

Then, by Equation (11),
\[ \text{Var}(\hat{\Delta}^{(ANCOVA)}) = \frac{1}{P(M_K = 1)} e_K^T C e_K = 27. \]

To compute \( \text{Var}(\hat{\Delta}^{(MMRM-II)}) \), recall that \( \nabla_M = V_M(\Sigma^{(MMRM-II)}) \). Then
\[ \nabla_{(1,0)} = \begin{pmatrix} \frac{1}{4} & 0 \\ 0 & 0 \end{pmatrix}, \quad \nabla_{(0,1)} = \begin{pmatrix} 0 & 0 \\ 0 & \frac{1}{4} \end{pmatrix}, \quad \nabla_{(1,1)} = \begin{pmatrix} -\frac{1}{15} & \frac{4}{15} \\ \frac{4}{15} & -\frac{1}{15} \end{pmatrix}, \]
which implies
\[ E^*[\nabla_M] = \sum_m p_m \nabla_m = \frac{1}{180} \begin{pmatrix} 31 & -4 \\ -4 & 31 \end{pmatrix}, \quad E^*[\nabla_M]^{-1} = \frac{4}{21} \begin{pmatrix} 31 & 4 \\ 4 & 31 \end{pmatrix}. \]

In addition, we have
\[ \nabla_{(1,0)} C \nabla_{(1,0)} = \begin{pmatrix} \frac{9}{8} & 0 \\ 0 & 0 \end{pmatrix}, \quad \nabla_{(0,1)} C \nabla_{(0,1)} = \begin{pmatrix} 0 & 0 \\ 0 & \frac{9}{8} \end{pmatrix}, \quad \nabla_{(1,1)} C \nabla_{(1,1)} = \frac{1}{50} \begin{pmatrix} 76 & -49 \\ -49 & 76 \end{pmatrix}, \]
which implies
\[ E^*[\nabla_M C \nabla_M] = \sum_m p_m \nabla_m C \nabla_m = \frac{1}{600} \begin{pmatrix} 529 & -196 \\ -196 & 529 \end{pmatrix}. \]

Then, by \( \Sigma_j^{(MMRM-II)} = \Sigma_j^{(ANCOVA)} \) and Equation (13), we have
\[ \text{Var}(\hat{\Delta}^{(MMRM-II)}) = e_K^T E^*[\nabla_M]^{-1} E^*[\nabla_M C \nabla_M] E^*[\nabla_M]^{-1} e_K \]
\[ = \frac{4}{21} \begin{pmatrix} 31 \\ -1 \end{pmatrix} \frac{1}{600} \begin{pmatrix} 529 & -196 \\ -196 & 529 \end{pmatrix} \frac{4}{21} \begin{pmatrix} 4 \\ -1 \end{pmatrix} \]
\[ = \frac{12486}{21^2} \approx 28.31. \]

Since \( 27 < 28.31 \), we have \( \text{Var}(\hat{\Delta}^{(ANCOVA)}) < \text{Var}(\hat{\Delta}^{(MMRM-II)}) \).
G Missing data mechanism in the simulation study

Given \((Y_i(0), Y_i(1), Y_i(2), A_i, X_{i1})\) defined in Section 6.1 of the main paper, we define \(R_{it}(j)\) as the residual of \(Y_{it}(j)\) regressing on \(X_{i1}\) by a simple linear regression. We then define the censoring time \(C_i\) by the following sequential conditional model

1. \(P(C_i = 1|X_{i1}, A_i) = 1 - \expit[\logit(0.99) - 0.14I\{A_i < 2\}X_{i1} - 0.12I\{A_i = 2\}X_{i1}]\),

2. \(P(C_i = 2|X_{i1}, A_i, C_i > 1, Y_{i1}) = 1 - \expit[\logit(0.969) - 0.7I\{A_i < 2\}R_{i1}(A_i) - 0.5I\{A_i = 2\}R_{i1}(A_i)]\),

3. \(P(C_i = 3|X_{i1}, A_i, C_i > 2, Y_{i1}, Y_{i2}) = 1 - \expit[\logit(0.958) - 0.72I\{A_i < 2\}R_{i2}(A_i) - 0.51I\{A_i = 2\}R_{i2}(A_i)]\),

4. \(P(C_i = 4|X_{i1}, A_i, C_i > 3, Y_{i1}, Y_{i2}, Y_{i3}) = 1 - \expit[\logit(0.967) - 0.74I\{A_i < 2\}R_{i3}(A_i) - 0.52I\{A_i = 2\}R_{i3}(A_i)]\),

5. \(P(C_i = 5|X_{i1}, A_i, C_i > 4, Y_{i1}, Y_{i2}, Y_{i3}, Y_{i4}) = 1 - \expit[\logit(0.977) - 0.76I\{A_i < 2\}R_{i4}(A_i) - 0.53I\{A_i = 2\}R_{i4}(A_i)]\).

Once we have \(C_i\), then \(M_i\) is defined as \(M_{it} = 1\) if \(t \leq C_i\) and 0 otherwise.

References

Bugni, F. A., Canay, I. A., and Shaikh, A. M. (2019). Inference under covariate-adaptive randomization with multiple treatments. *Quantitative Economics*, 10(4):1747–1785.

Dwyer, P. S. (1967). Some applications of matrix derivatives in multivariate analysis. *Journal of the American Statistical Association*, 62(318):607–625.
MacRae, E. C. et al. (1974). Matrix derivatives with an application to an adaptive linear decision problem. *The Annals of Statistics, 2*(2):337–346.

van der Vaart, A. (1998). *Asymptotic Statistics*. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press.

Wang, B., Susukida, R., Mojtahai, R., Amin-Esmaeili, M., and Rosenblum, M. (2021). Model-robust inference for clinical trials that improve precision by stratified randomization and covariate adjustment. *Journal of the American Statistical Association*, pages 1–12.