
REPORT ON MODEL-ROBUST INFERENCE FOR CLINICAL TRIALS THAT IMPROVE PRECISION BY STRATIFIED RANDOMIZATION AND COVARIATE ADJUSTMENT

Zhihan Huang, Ruizhe Deng

June 24, 2022

1 Introduction

A joint guidance document from the U.S. Food and Drug Administration and the European Medicines Agency (FDA and EMA 1998[1]) states that “Pretrial deliberations should identify those covariates and factors expected to have an important influence on the primary variable(s), and should consider how to account for these in the analysis to improve precision and to compensate for any lack of balance between treatment groups.” Many recent regulatory guidance documents encourage consideration of baseline variables in order to improve precision in randomized trials. As the most widely used method, simple randomization has been discussed by a rich literature and many statistical methods have been used to improve precision. However, there is a less concern about other forms of randomization, which is also of significant practical use.

“Covariate-adaptive randomization” refers to randomization procedures that take baseline variables into account when assigning participants to study arms, which we mainly concern in this article. By contrast, the simple randomization ignores baseline variable and design assignment mechanism based on the whole population. Thus, covariate-adaptive randomization can be advantageous in minimizing imbalance and improving efficiency and has become a widely used approach in clinical trials. According to a survey by Lin, Zhu, and Su (2015)[2], 183 out of their sample of 224 randomized clinical trials published in 2014 in leading medical journals used some form of covariate-adaptive randomization.

We only consider the following two types of covariate-adaptive randomization since they are the most frequently implemented methods in clinical trials: stratified randomization and biased-coin randomization. Specific procedure and details about these two randomization methods will be given in the following part.

In the preceding statistical analyses at the end of trials, adjusting for stratification variables is recommended by literature for its advantages over simple randomization. However, there exist two main problems in the previous study.

First, most of the studies do not reliably carry out theoretical reasons for the recommendation. According to Kahan and Morris (2012)[3], among 65 published trials from major medical journals from March to May 2010, 41 of them implemented covariate-adaptive randomization (among which 29 used stratified randomization), but only 14 adjusted in the primary analysis for the variables used in the randomization procedure.

Second, many results on how to conduct the primary efficacy analysis in trials that use stratified randomization require extra assumption about the model. The theorem in this article extend the conclusion when there is random missing data or existing missing mechanism.

For trials using stratified or biased-coin randomization, it was an open problem to determine the large sample properties of causal estimators. Under regularity conditions, the writers prove that a large class of estimators is consistent and asymptotically normally distributed in randomized trials that use stratified or biased-coin randomization, and extend the formula for computing their asymptotic variance in simple randomization to other randomization methods. This class of estimators consists of all regular M-estimators for binary and continuous outcomes and Kaplan–Meier (K-M) estimator (Kaplan and Meier 1958[4]) for time-to-event outcomes. In brief, the theorems imply that under standard regularity conditions, whenever an estimator in our class is consistent and asymptotically normally distributed under

simple randomization, then it is consistent and asymptotically normally distributed under stratified (or biased-coin) randomization. Also, the new asymptotic variance function can be calculated by a simple transformation.

Simulations and clinical trial analyses are made in the later section. As we show in our data analyses, the addition of baseline variables beyond those used for stratified randomization can lead to substantial precision gains.

In this article, the commonly used superpopulation inference framework is applied in estimation and no parametric assumptions are made before analysis. As in the aforementioned work, we assume that the randomization procedure and analysis method have been completely specified before the trial starts, as is typically required by regulators.

2 Randomization Procedures

In this section, we will specify the randomization procedures concerned in the article including simple randomization, stratified randomization and biased coin randomization, then briefly explain their properties.

First consider simple randomization, which assigns study arms A_1, \dots, A_n by independent Bernoulli draws each with fixed probability π of being 1, for example, using a random number generator. By design, the draws are independent of each other and of all participant characteristics measured before randomization or not impacted by randomization. Therefore, we have that (A_1, \dots, A_n) is independent of the full data of participants (W_1, \dots, W_n) , and also, under assumptions below, the observed data O_1, \dots, O_n are independent, identically distributed. It is the most commonly used method in practical experiments and there are many results about estimators under it in the preceding literature.

Next consider stratified or biased-coin randomization, where treatment allocation depends on predefined baseline strata, such as gender, age, site, disease severity, or combinations of these. In the article, we refer to the baseline strata that are used in the randomization procedure as “randomization strata.” The baseline stratum of participant i is denoted by a categorical variable S_i taken value in the set $S = \{1, \dots, K\}$ denotes all K randomization strata. In stratified or biased-coin randomization, we try to achieve balance in each stratum instead of in the whole population; that is, the proportion of participants assigned to the treatment arm in each stratum is targeted to the prespecified proportion $\pi \in (0, 1)$.

We use “stratified permuted block randomization” in Zelen (1974)[5] as the definition of stratified randomization. It uses permuted blocks to assign treatment. For each randomization stratum, a randomly permuted block with fraction π 1’s (representing treatment) and $(1 - \pi)$ 0’s (representing control) is used for sequential allocation. When a block is exhausted, a new block is used. This procedure repeats until completing the assignment of one stratum, then turn to the next stratum. When the number of participants in each stratum is large enough, the balance is approximately achieved.

Biased-coin randomization was defined in Efron (1971)[6]. It can be applied when $\pi = 0.5$ and it allocates participants sequentially by the following rule for $k = 1, \dots, n$:

$$P(A_k = 1 | S_1, \dots, S_k, A_1, \dots, A_{k-1}) = \begin{cases} 0.5, & \text{if } \sum_{i=1}^{k-1} (A_i - 0.5) I\{S_i = S_k\} = 0 \\ \lambda, & \text{if } \sum_{i=1}^{k-1} (A_i - 0.5) I\{S_i = S_k\} < 0 \\ 1 - \lambda, & \text{if } \sum_{i=1}^{k-1} (A_i - 0.5) I\{S_i = S_k\} > 0, \end{cases}$$

where $\lambda \in (0.5, 1]$, I_Z is the indicator function and by convention the first participant is assigned with probability 0.5 to each arm. Our results for biased-coin randomization assume that $\pi = 0.5$. Intuitively, for each stratum, when a new participant joining, we compare the number of participants in treatment and control group. If there are more people in the treatment group, the new one will be assigned to control group with a higher probability. Otherwise, the probability of being assigned to treatment group is higher.

3 Proof

The main conclusions of the article focus on two kinds of outcomes: the real-value outcomes and the time-to-events outcomes. Theorem 1 shows the properties of M-estimator of real-value outcomes under simple randomization and stratified randomization (or biased-coin). Theorem 2 extends the conclusions to time-to-event data and commonly used Kaplan-Meier estimator under simple randomization and stratified randomization (or biased-coin).

3.1 Real-Value Outcome

First, we define the full data vector $W_i = (Y_i(0), Y_i(1), M_i(0), M_i(1), X_i)$ and observed data vector $O_i = (A_i, X_i, Y_i M_i, M_i)$. The product $Y_i M_i$ is used in the data to encode that whenever the outcome is missing ($M_i = 0$),

the outcome value Y_i is not available in O_i . We use E to denote expectation with respect to full data distribution P and E^* to denote expectation with respect to observed data distribution P^* . Consistency assumption is always made that $Y = AY(1) + (1 - A)Y(0)$ and $M = AM(1) + (1 - A)M(0)$.

For real-value outcomes, like binary and continuous outcomes, our goal is to estimate a population parameter Δ^* , which is a contrast between the marginal distribution of $Y(1)$ and $Y(0)$. For example, Δ^* can be the average treatment effect $E(Y(1) - Y(0))$. Here we consider M-estimators of Δ^* . We define the M-estimator $\hat{\theta} = (\hat{\Delta}, \hat{\beta})$ to be solution of estimating equation:

$$\sum_{i=1}^n \psi(A_i, X_i, Y_i, M_i; \theta) = 0, \quad (1)$$

where $\Delta \in \mathbb{R}$ is the parameter of interest and $\beta \in \mathbb{R}^p$ is a column vector of nuisance parameters. We assume that ψ does not depend on the outcome Y when $M = 0$ (since it implies that Y is missing).

For real-value outcomes, we make the following assumptions on the distribution of the full data vector $\{W_1, W_2, \dots, W_n\}$: **Assumption 1.** (1) $W_i, i = 1, 2, \dots, n$ are i.i.d. samples from an unknown joint distribution P .

(2) Missing at random: $M(a) \perp\!\!\!\perp Y(a) \mid X$ for each arm $a \in \{0, 1\}$.

Here we make the formal statement of Theorem 1.

Theorem 1. Assume the regular conditions of the estimating equations. We use $\hat{\Delta}$ to denote the M-estimator of Δ^* . Then under simple, stratified, or biased-coin randomization, we have consistency, that is $\hat{\Delta} \rightarrow \Delta^*$ in probability and asymptotic linearity, that is,

$$\sqrt{n}(\hat{\Delta} - \Delta^*) = \frac{1}{\sqrt{n}} \sum_{i=1}^n IF(A_i, X_i, Y_i, M_i) + o_p(1) \quad (2)$$

where the influence function $IF(A, X, Y, M)$ is the first entry of $-B^{-1}\psi(A, X, Y, M; \theta^*)$ for $B = E^*\left[\frac{\partial}{\partial \theta}\psi(A, X, Y, M; \theta)\right]_{\theta=\theta^*}$.

For stratified and biased-coin randomization, $\sqrt{n}(\hat{\Delta} - \Delta^*) \xrightarrow{d} N(0, V)$ for

$$V = \tilde{V} - \frac{1}{\pi(1 - \pi)} E^*[E^*[(A - \pi)IF(A, X, Y, M)|S]^2], \quad (3)$$

where $\tilde{V} = E^*[IF(A, X, Y, M)^2]$ is the asymptotic variance under simple randomization.

Proof. Under simple randomization, the results of Theorem 1 have been proved in Section 5 of van der Vaart (1998)[7], Theorem 5.41, which holds for independent, identically distributed data. Use Taylor expansion of the function ψ at the neighborhood of $\hat{\Delta}$.

Hence, it remains to prove Theorem 1 under stratified or biased-coin randomization, which we do below. The important difference between simple randomization and stratified randomization is that in stratified randomization, the study arm assignment of individual i , which is A_i is dependent of the former individuals. Since in each strata, the individuals are assigned to treatment or control group one by one through a certain mechanism, the distribution of A_i is affected by the former individuals. However, in each strata, the study arm assignment is independent of the full data vector W_i . We have

$$(A_1, \dots, A_n) \perp\!\!\!\perp (W_1, \dots, W_n) \mid (S_1, \dots, S_n). \quad (4)$$

We can use this conclusion to derive properties of assignment in stratified randomization.

Properties. (1) For each strata s , fix the strata assignment (S_1, \dots, S_n) , we have

$$\frac{1}{\sqrt{n}} \sum_{i=1}^n (A_i - \pi) I\{S_i = s\} \xrightarrow{P} 0. \quad (5)$$

(2) Under stratified randomization, each observed data $O_i = (A_i, X_i, Y_i, M_i)$ are identically distributed and A_i is independent of W_i , for $i = 1, \dots, n$, which means $P(A_i = 1) = \pi$ still holds.

The properties shows that for each individual, the stratified randomization is the same as the simple randomization and when the sample size n is large enough, the distributions of study arm assignment (A_1, \dots, A_n) in simple and stratified randomization tends to be the same.

To simplify the notation, we define $\psi^{(a)}(\theta) = \psi(a, X, Y(a), M(A); \theta)$ and $\psi_i^{(a)}(\theta) = \psi(a, X_i, Y_i(a), M_i(A); \theta)$ for $a = 0, 1$ and $i = 1, \dots, n$. Use the fact of consistency:

$$\begin{aligned} Y_i &= Y_i(1)A_i + Y_i(0)(1 - A_i), \\ M_i &= M_i(1)A_i + M_i(0)(1 - A_i), \end{aligned}$$

the estimating equations can be re-written as

$$\frac{1}{n} \sum_{i=1}^n (A_i \psi_i^{(1)}(\theta) + (1 - A_i) \psi_i^{(0)}(\theta)) = \mathbf{0},$$

and the true value of the model parameter θ^* is assumed to be the solution of

$$\pi E[\psi_i^{(1)}(\theta)] + (1 - \pi) E[\psi_i^{(0)}(\theta)] = \mathbf{0}.$$

Like the simple randomization scenario, we apply multivariate Taylor expansion on the function $\sum_{i=1}^n (A_i \psi_i^{(1)}(\theta) + (1 - A_i) \psi_i^{(0)}(\theta))$ around the point $\theta = \theta^*$:

$$\begin{aligned} \mathbf{0} &= \frac{1}{n} \sum_{i=1}^n (A_i \psi_i^{(1)}(\hat{\theta}) + (1 - A_i) \psi_i^{(0)}(\hat{\theta})) \\ &= \frac{1}{n} \sum_{i=1}^n (A_i \dot{\psi}_i^{(1)}(\theta^*) + (1 - A_i) \dot{\psi}_i^{(0)}(\theta^*)) + \frac{1}{n} \sum_{i=1}^n (A_i \ddot{\psi}_i^{(1)}(\tilde{\theta}) + (1 - A_i) \ddot{\psi}_i^{(0)}(\tilde{\theta}))(\hat{\theta} - \theta^*) \\ &\quad + \frac{1}{2}(\hat{\theta} - \theta^*)^t \frac{1}{n} \sum_{i=1}^n (A_i \ddot{\psi}_i^{(1)}(\tilde{\theta}) + (1 - A_i) \ddot{\psi}_i^{(0)}(\tilde{\theta}))(\hat{\theta} - \theta^*). \end{aligned}$$

According to regularity conditions, $\|\ddot{\psi}_i^{(a)}(\theta)\|$ is dominated by an integrable function $v(X_i, Y_i(a), M_i(a))$. Thus, we have

$$\left\| \frac{1}{n} \sum_{i=1}^n (A_i \ddot{\psi}_i^{(1)}(\tilde{\theta}) + (1 - A_i) \ddot{\psi}_i^{(0)}(\tilde{\theta})) \right\| = O_p(1).$$

Then using the conditional independence of assignment mechanism and other factors, we can show that

$$\frac{1}{n} \sum_{i=1}^n (A_i \dot{\psi}_i^{(1)}(\theta^*) + (1 - A_i) \dot{\psi}_i^{(0)}(\theta^*)) \xrightarrow{P} \pi E[\dot{\psi}^{(1)}(\theta)] + (1 - \pi) E[\dot{\psi}^{(0)}(\theta)]. \quad (6)$$

Recalling that we set

$$\begin{aligned} B &= E^* \left[\frac{\partial}{\partial \theta} \psi(A, X, Y, M; \theta) \Big|_{\theta=\theta^*} \right] \\ &= E^*(E^*[\dot{\psi}(\theta^*)|A]) \\ &= \pi E[\dot{\psi}^{(1)}(\theta)] + (1 - \pi) E[\dot{\psi}^{(0)}(\theta)], \end{aligned}$$

the equation (6) can be tr-written as

$$\frac{1}{n} \sum_{i=1}^n (A_i \dot{\psi}_i^{(1)}(\theta^*) + (1 - A_i) \dot{\psi}_i^{(0)}(\theta^*)) = B + o_p(1).$$

After calculate the remainder, we get

$$\sqrt{n}(\hat{\theta} - \theta^*) = -B^{-1} \frac{1}{\sqrt{n}} \sum_{i=1}^n \psi(A_i, X_i, Y_i, M_i; \theta^*) + o_p(1).$$

Replace θ with a specific parameter Δ and we complete the proof of asymptotic linearity.

Then we turn to the asymptotic variance. Under simple randomization, using center limit theorem and the fact that $A_i, i = 1, \dots, n$ are independent, we get

$$\begin{aligned} \text{Var}(\sqrt{n}(\hat{\Delta} - \Delta^*)) &\rightarrow \frac{1}{n} \sum_{i=1}^n \text{Var}(IF(A_i, X_i, Y_i, M_i)) \\ &= E^*[IF(A, X, Y, M)^2] = \widetilde{V}. \end{aligned}$$

Under stratified randomization, the asymptotic variance has a different form because of the dependence of influence functions. To simplify the notation, we denote $IF(a) = IF(a, X, Y(a), M(a))$ and $IF_i(a) = IF(a, X_i, Y_i(a), M_i(a))$ for $a = 0, 1$ and $i = 1, \dots, n$. Using property (1), the equation (2) can be re-written as

$$\sqrt{n}(\hat{\Delta} - \Delta^*) = \frac{1}{\sqrt{n}} \sum_{i=1}^n [A_i(IF_i(1) - E[IF(1)]) + (1 - A_i)(IF_i(0) - E[IF(0)])] + o_p(1).$$

We assume that $\sqrt{n}(\hat{\Delta} - \Delta^*) \xrightarrow{d} N(0, V)$. Using conditional variance formula: $Var(Y) = Var(E(Y|X)) + E(Var(Y|X))$, we have the expression

$$\begin{aligned} V &= \pi Var(IF(1) - E[IF(1)|S]) + (1 - \pi) Var(IF(0) - E[IF(0)|S]) \\ &\quad + Var(E[\pi IF(1) + (1 - \pi) IF(0)|S]) \\ &= \pi E[IF(1)^2] + (1 - \pi) E[IF(0)^2] - \pi(1 - \pi) E[E[IF(1) - IF(0)|S]^2] \\ &= \tilde{V} - \pi(1 - \pi) E[E[IF(1) - IF(0)|S]^2], \end{aligned}$$

and that

$$\begin{aligned} &E^* \left[\frac{A - \pi}{\pi(1 - \pi)} IF(A, X, Y, M) | S \right] \\ &= E^* \left[E^* \left[\frac{A - \pi}{\pi(1 - \pi)} IF(A, X, Y, M) | A \right] | S \right] \\ &= E[IF(1) - IF(0) | S]. \end{aligned}$$

Combining all the conclusions above, we have

$$V = \tilde{V} - \frac{1}{\pi(1 - \pi)} E^* [E^* [(A - \pi) IF(A, X, Y, M) | S]^2]. \quad (7)$$

Theorem 1 implies that whenever an M-estimator $\hat{\Delta}$ is consistent and asymptotically normally distributed under simple randomization, then it is consistent and asymptotically normally distributed under stratified (or biased-coin) randomization with equal or smaller asymptotic variance, which is more powerful in estimation. Also, its influence function is the same regardless of whether data are generated under simple, stratified, or biased-coin randomization.

In order to construct confidence intervals (CIs) and perform hypothesis tests, one can use the following estimator for the asymptotic variance \tilde{V} and V .

For simple randomization and corresponding M-estimator, since Δ is the first component of model parameter θ , we can use the sandwich estimator \tilde{V}_n , defined as the first entry of

$$\tilde{V}_n = \left\{ E_n \left[\frac{\partial \psi}{\partial \theta} \middle|_{\theta = \hat{\theta}} \right] \right\}^{-1} \left\{ E_n [\psi \psi^t \middle|_{\theta = \hat{\theta}}] \right\} \left\{ E_n \left[\frac{\partial \psi}{\partial \theta} \middle|_{\theta = \hat{\theta}} \right] \right\}^{-1, t},$$

where E_n denotes expectation with respect to the empirical distribution of the observed data.

For stratified (or biased-coin) randomization, equation (7) can be used to construct variance estimator \hat{V} :

$$\hat{V} = \tilde{V}_n - \frac{1}{\pi(1 - \pi)} E_n [E_n [(A - \pi) IF(A, X, Y, M) | S]^2].$$

Then a CI for Δ^* can be constructed based on the normal approximation with variance \hat{V}/n , which we used in our simulations and clinical trial study.

Many estimators used in clinical trials can be expressed as solutions to estimating equations (1) for an appropriately chosen estimating function ψ .

For example, for binary outcome, the commonly used estimator is standardized logistic regression estimator: Δ_{log} . We assume that there is no missing data in this model. It is calculate by first fitting model $P(Y = 1 | A, X) = \text{expit}(\beta_0 + \beta_A A + \beta_X^t X)$, where $\text{expit}(x) = 1/(1 + e^{-x})$, and getting the ML estimates $(\hat{\beta}_0, \hat{\beta}_A, \hat{\beta}_X)$. Then the logistic regression estimator Δ_{log} is defined as the solution to estimating equations:

$$\psi(A, X, Y, M; \theta) = \begin{pmatrix} \text{expit}(\beta_0 + \beta_A A + \beta_X^t X) - \text{expit}(\beta_0 + \beta_X^t X) - \Delta \\ (Y - \text{expit}(\beta_0 + \beta_A A + \beta_X^t X))Z \end{pmatrix},$$

where $Z = (1, A, X^t)^t$. It is equivalent to the common definition that $\Delta_{log} = \frac{1}{n} \sum_{i=1}^n (\text{expit}(\hat{\beta}_0 + \hat{\beta}_A A + \hat{\beta}_X^t X) - \text{expit}(\hat{\beta}_0 + \hat{\beta}_X^t X))$.

When some outcomes are missing and the missing at random assumption holds, also the outcomes can be either continuous or binary, one extension is DR-WLS estimator. This estimator is calculated by first fitting the logistic regression model for missing mechanism:

$$P(M = 1|A, X) = \text{expit}(\alpha_0 + \alpha_A A + \alpha_X^t X) \quad (8)$$

and getting the ML estimates $(\hat{\alpha}_0, \hat{\alpha}_A, \hat{\alpha}_X^t)$. Next, fitting the generalized linear model(GLM):

$$E[Y|A, X] = g^{-1}(\beta_0 + \beta_A A + \beta_X^t X), \quad (9)$$

with weights $1/\text{expit}(\hat{\alpha}_0 + \hat{\alpha}_A A_i + \hat{\alpha}_X^t X_i)$ for individual i . Here g^{-1} is the inverse link function of the model, which can be chosen appropriately for different outcomes. For example, $g^{-1}(x) = \text{expit}$ for binary outcomes and $g^{-1}(x) = x$ for continuous outcomes. While fitting the model, we can only use the data with $M_i = 1$ (non-missing), so it shares the same idea with inverse propensity score(IPW) estimator[8].

Combining the procedure above, the DR-WLS estimator can be written as the solution to the estimating equations:

$$\psi(A, X, Y, M; \theta) = \begin{pmatrix} \frac{g^{-1}(\beta_0 + \beta_A A + \beta_X^t X) - g^{-1}(\beta_0 + \beta_X^t X) - \Delta}{\frac{M}{\text{expit}(\hat{\beta}_0 + \hat{\beta}_A A + \hat{\beta}_X^t X)}} (Y - g^{-1}(\beta_0 + \beta_A A + \beta_X^t X)) Z \\ (M - \text{expit}(\alpha_0 + \alpha_A A + \alpha_X^t X)) Z \end{pmatrix}.$$

It is equivalent to the direct definition that

$$\hat{\Delta}_{DR-WLS} = \frac{1}{n} \sum_{i=1}^n (g^{-1}(\beta_0 + \beta_A A + \beta_X^t X) - g^{-1}(\beta_0 + \beta_X^t X)).$$

The DR-WLS estimator can be generalized to allow the addition of interaction terms in the model, while it can still be treated as a regular M-estimator in the same way.

The DR-WLS estimator is widely used since some simple estimators can be regarded as special cases of it, for example ANCOVA estimator and logistic regression estimator. As an extension of these robust estimator, the DR-WLS estimator is doubly robust estimator, that is, if we take the overlap assumption for missing mechanism:

$$\inf_{(a,x) \in (\mathcal{A}, \mathcal{X})} P(M = 1|a, x) > 0,$$

where $(\mathcal{A}, \mathcal{X})$ is the support of (A, X) , then we have the estimator is consistent and asymptotically normal if at least one of model (8),(9) is correctly specified.

For the DR-WLS estimator, the expression (3) can be used to calculate the asymptotic variance under stratified and biased-coin randomization. However, asymptotic variance will not necessarily decrease because of stratification. The influence function in the expression can be calculate from Theorem 5.21 of van der Vaart (1998). We have

$$\tilde{V} - V = \frac{(1 - 2\pi)}{\pi^3(1 - \pi)^3} E^* \left[E^* \left[\frac{AM(Y - h(A, X))}{e(A, X)} \middle| S \right]^2 \right], \quad (10)$$

where $h(A, X) = g^{-1}(\underline{\beta}_0 + \underline{\beta}_A A + \underline{\beta}_X^t X)$ and $e(A, X) = \text{expit}(\underline{\alpha}_0 + \underline{\alpha}_A A + \underline{\alpha}_X^t X)$ and $\underline{\alpha}_0, \underline{\alpha}_A, \underline{\alpha}_X, \underline{\beta}_0, \underline{\beta}_A, \underline{\beta}_X$ are the probability limits of the ML estimator. Use equation (10), we can find some sufficient conditions for $\tilde{V} - V = 0$.

Corollary 1. If any of the conditions (a)–(c) below holds, then under simple, stratified, or biased-coin randomization, the estimator is consistent and asymptotically normally distributed with asymptotic variance $\tilde{V} = V$; furthermore, the sandwich variance estimator is consistent. Conditions:

- (a) $\pi = 0.5$;
- (b) the outcome regression model includes indicators for the randomization strata and also treatment-by-randomization-strata interaction terms;
- (c) the outcome regression model is correctly specified.

It reveals that in many cases, the stratified method will not improve the asymptotic power of DR-WLS estimator, so sandwich estimator can be used in all these randomization methods. But for finite sample data analysis, adjusted estimator may still show some superiority. See real data analysis.

3.2 Time-to-Event Outcome

For time to event outcomes, since we need to estimate a survival function rather than a real number or a vector, the theorem and proof above do not apply. Thus, we use slightly modified notation and assumptions compared to above.

As in most cases, the outcomes is assumed to be right-censored, which means that the true survival times will always be equal to or greater than the observed survival time. We use notation analogous as the system in Theorem 1. Let Y_i denote the failure time and M_i denote the censoring time of individual i . Other variables including A_i, X_i and the potential outcomes $Y_i(a), M_i(a)$ for $a = 0, 1$ are defined analogously as before.

For each participant $i \in \{1, \dots, n\}$, we define the full data vector $W_i = (Y_i(0), Y_i(1), M_i(0), M_i(1), X_i)$ and observed data vector $O_i = (A_i, X_i, U_i, \delta_i)$, where $U_i = \min\{Y_i, M_i\}$ and $\delta_i = I\{Y_i < M_i\}$. U_i is the observed time and δ_i is the indicator of missing data (when $\delta_i = 1$, the event is observed and when $\delta_i = 0$, censoring happens and the data is missing). We further define a restriction time τ such that the time window $t \in [0, \tau]$ is of interest. We use E to denote expectation with respect to full data distribution P and E^* to denote expectation with respect to observed data distribution P^* .

The goal of our analysis is to estimate the survival function curve $\{S_0^{(a)}(t) : t \in [0, \tau]\}$ for each $a = 0, 1$, where the survival function in this causal study is defined as $S_0^{(a)}(t) = P(Y(a) > t)$. This represents the survival curve if everyone in the study population were assigned to study arm a .

Here we consider the non-parametric estimators of survival functions. One commonly used method for survival analysis is the Kaplan-Meier estimator $\hat{S}_n^{(a)}(t)$:

$$\hat{S}_n^{(a)}(t) = \prod_{j: T_j \leq t} \left(1 - \frac{\sum_{i=1}^n \delta_i I\{A_i = 1\} I\{U_i = T_j\}}{\sum_{i=1}^n I\{A_i = 1\} I\{U_i \geq T_j\}}\right) \quad (11)$$

Unlike M-estimator we consider above, the K-M estimator does not adjust for any baseline variables. Thus its variance under simple randomization is typically different than under stratified or biased-coin randomization, and this is not accounted for by standard methods for estimating its variance. Specifically, the standard method for variance estimation will typically overestimate the K-M variance under stratified or biased-coin randomization, leading to wasted power.

Here we will use Theorem to show the properties of K-M estimator under stratified (or biased-coin) randomization compared with simple randomization. The following assumption is made in place of Assumption 1:

Assumption 2.: (1) $W_i, i = 1, 2, \dots, n$ are i.i.d. samples from an unknown joint distribution P .

(2) Censoring at random: $M(a) \perp\!\!\!\perp Y(a)$ for each arm $a \in \{0, 1\}$.

(3) $P(\min\{Y(a), M(a)\} > \tau) > 0$ for each $a = 0, 1$.

Compared with Assumption 1, Assumption 2(1) is the same as Assumption 1(1), and Assumption 2(2) assumes censoring completely at random instead of missing at random. This modification of the assumption on missing data is to adjust the K-M estimator and its consistency requirement. Assumption 2(3) is often made in survival analysis, which states that there is a positive probability that both the failure time and censoring time occur after τ (under each study arm assignment).

Here we make the formal statement of Theorem 2.

Theorem 2. Given Assumption 2, under simple, stratified, or biased-coin randomization, we have for each $t \in [0, \tau]$, $a = 0, 1$ that

$$\sqrt{n}(\hat{S}_n^{(a)}(t) - S_0^{(a)}(t)) = \frac{1}{\sqrt{n}} \sum_{i=1}^n IF^{(a)}(A_i, U_i, \delta_i; t) + o_{p^*}(1) \quad (12)$$

where $o_{p^*}(1)$ represents a sequence of random variables converging to 0 in probability uniformly over $t \in [0, \tau]$ and $IF^{(a)}(A_i, U_i, \delta_i; t)$ is the corresponding influence function which will be given in the proof.

For stratified and biased-coin randomization, the random process $\{\sqrt{n}(\hat{S}_n^{(a)}(t) - S_0^{(a)}(t)) : t \in [0, \tau]\}$ converges weakly to a mean 0, tight Gaussian process with covariance function $V^{(a)}(t, t')$, which has the following property: for any $t \leq \tau$,

$$V^{(a)}(t, t) = \widetilde{V}^{(a)}(t, t) - \frac{1}{\pi(1 - \pi)} E^*[E^*[(A - \pi)IF^{(a)}(A, U, \delta; t)|S]^2], \quad (13)$$

where $\widetilde{V}^{(a)}(t, t')$ is the asymptotic variance under simple randomization. The concrete form of $V^{(a)}(t, t')$ will be given in the proof.

Proof. First we introduce notations used in survival analysis.

For each $a = 0, 1$ and $t \in [0, \tau]$, define counting process $N(t) = I\{U \leq t, \delta = 1\}$ and $N^{(a)}(t) = I\{U(a) \leq t, \delta(a) = 1\}$. Then the cumulative hazard function $\Lambda^{(a)}(t)$:

$$\Lambda^{(a)}(t) = E\left[\int_0^t \frac{dN^{(a)}(t')}{P(U(a) \geq t')}\right],$$

and the martingale $L^{(a)}(t)$:

$$L^{(a)}(t) = N^{(a)}(t) - \int_0^t I\{U(a) \geq t'\} d\Lambda^{(a)}(t'),$$

the martingale transformation $H^{(a)}(t)$:

$$H^{(a)}(t) = \int_0^t \frac{dL^{(a)}(t')}{(1 - \Delta\Lambda^{(a)}(t'))P(U(a) \geq t')},$$

where $\Delta\Lambda^{(a)}(t) = \Lambda^{(a)}(t) - \Lambda^{(a)}(t-)$ accounting for the discontinuous point. Then we define the corresponding non-parametric estimator of these function:

$$\hat{\Lambda}^{(a)}(t) = \int_0^t \frac{\sum_{i=1}^n I\{A_i = a\} dN_i(t')}{\sum_{i=1}^n I\{A_i = a\} I\{U_i \geq t'\}} \quad (14)$$

$$\Delta\hat{\Lambda}^{(a)}(t) = \frac{\sum_{i=1}^n I\{A_i = a\} I\{U_i = t, \delta_i = 1\}}{\sum_{i=1}^n I\{A_i = a\} I\{U_i \geq t\}}, \quad (15)$$

which we get from replacing the expectation with respect to true distribution by the empirical distribution \mathbb{P}_n .

For simple randomization, it is a classical properties of K-M estimator. We refer to some textbooks, Chapter 3 of Fleming and Harrington (1991)[9] and Chapter 4 of Kosorok (2008)[10].

For stratified randomization, we first prove that $\hat{\Lambda}^{(a)}(t)$ is asymptotically linear. Re-writing the expression of $\hat{\Lambda}^{(a)}(t)$:

$$\begin{aligned} & \hat{\Lambda}^{(a)}(t) - \Lambda^{(a)}(t) \\ &= \int_0^t \frac{\mathbb{P}_n I\{A_i = a\} dL^{(a)}(t')}{\mathbb{P}_n I\{A_i = a\} I\{U(a) \geq t'\}} - \int_0^t I\{\mathbb{P}_n I\{A = a\} I\{U(a) \geq t'\} = 0\} d\Lambda^{(a)}(t') \\ &= \frac{1}{n} \sum_{i=1}^n \frac{I\{A_i = a\}}{a\pi + (1-a)(1-\pi)} \int_0^t \frac{dL_i^{(a)}(t')}{P(U(a) \geq t')} + I_n(t). \end{aligned}$$

Then bounding the integrand of the remainder with $o_p^*(1/\sqrt{n})$ and using the fact from Assumption 2(3) that $\Lambda^{(a)}(t)$ is bounded, we can prove that

$$I_n(t) \leq o_p^*(1/\sqrt{n})\Lambda^{(a)}(\tau) = o_p^*(1/\sqrt{n}),$$

which implies the asymptotic linearity of $\hat{\Lambda}^{(a)}(t)$.

Define a Hadamard differentiable map $\psi(f)(t) = \prod_{0 \leq t' \leq t} (1 + df(t'))$. Then use the relationship between hazard function and survival function, we can show that $\sqrt{n}(\hat{S}_n^{(a)}(t) - S_0^{(a)}(t)) = \psi_{-\Lambda}'(-\sqrt{n}(\hat{\Lambda}^{(a)}(t) - \Lambda^{(a)}(t))) + o_p^*(1)$. Through the calculation of ψ' , we have the expression:

$$\sqrt{n}(\hat{S}_n^{(a)}(t) - S_0^{(a)}(t)) = -\frac{a}{\sqrt{n}} \sum_{i=1}^n \frac{I\{A = a\}}{a\pi + (1-a)(1-\pi)} S_0^{(a)}(t) H^{(a)}(t) + o_p^*(1),$$

which proves the asymptotic linearity claimed in Theorem 2 and shows that

$$IF^{(a)}(A, U, \delta; t) = -\frac{I\{A = a\}}{a\pi + (1-a)(1-\pi)} S_0^{(a)}(t) H^{(a)}(t). \quad (16)$$

In application, we mainly focus on the asymptotic variance function of survival function at one fixed point to construct confidence intervals. For simple randomization, this function has been calculate in Zhang (2015):

$$\tilde{V}^{(a)}(t, t) = \frac{S_0^{(a)}(t)^2}{a\pi + (1-a)(1-\pi)} \text{Var}(H^{(a)}(t)).$$

Follow use the same technique in Theorem 1, we have

$$\begin{aligned} \tilde{V}^{(a)}(t, t) - V^{(a)}(t, t) &= \frac{1}{\pi(1-\pi)} E^*[E^*[(A-\pi)IF^{(a)}(A, U, \delta; t)|S]^2] \\ &\geq 0, \end{aligned} \quad (17)$$

which implies that stratified randomization will give a narrower confidence band in estimation.

From the point view of random processes, $\tilde{V}^{(a)}(t, t) - V^{(a)}(t, t)$ is in proportion to sum of the product of the martingale $H^{(a)}(t)$ and the known term $S_0^{(a)}(t)$. Thus, the properties of stratified randomization can be used to treat the dependence term and then use the martingale central limit theorem to get the limit process and show it is a tight Gaussian process. The details of the usage of martingale central limit theorem are shown in Bugni (2018)[11] and (2019)[12]. The covariance function of the tight Gaussian process are given below:

$$V^{(a)}(t, t') = S_0^{(a)}(t)S_0^{(a)}(t') \left\{ \frac{1}{a\pi + (1-a)(1-\pi)} E[\text{Cov}(H^{(a)}(t), H^{(a)}(t')|S)] \right. \quad (18)$$

$$\left. + \text{Cov}(E[H^{(a)}(t)|S], E[H^{(a)}(t')|S]) \right\}. \quad (19)$$

In real data analysis, we need to estimate the covariate function from the data, especially the variance function $V^{(a)}(t, t)$ to obtain the confidence band. Like the real-value outcome senario, we use the plug-in estimator

$$\hat{V}^{(a)}(t, t) = \frac{\hat{S}_n^{(a)}(t)^2}{\pi_a} \left[\hat{B}^{(a)}(t) - (1 - \pi_a) \sum_{s \in \mathcal{S}} \hat{p}(s) \left(\frac{\sum_{i=1}^n I\{S_i = s\} \hat{H}_i^{(a)}(t)}{\sum_{i=1}^n I\{S_i = s\}} \right)^2 \right],$$

where $\hat{S}_n^{(a)}(t)$ is the Kaplan-Meier estimator defined above and

$$\begin{aligned} \pi_a &= \pi a + (1-\pi)(1-a), \\ \hat{B}^{(a)}(t) &= \int_0^t \frac{d\hat{\Lambda}^{(a)}(t')}{\hat{P}(U(a) \geq t') \{1 - \Delta \hat{\Lambda}^{(a)}(t')\}}, \\ \hat{H}_i^{(a)}(t) &= \int_0^t \frac{I\{A_i = a\} (dN_i(t') d\hat{\Lambda}^{(a)}(t')) - I\{U_i \geq t'\}}{\pi_a \hat{P}(U(a) \geq t') \{1 - \Delta \hat{\Lambda}^{(a)}(t')\}}, \end{aligned}$$

with

$$\hat{P}(U(a) \geq t) = \frac{\sum_{i=1}^n I\{A_i = a\} I\{U_i \geq t\}}{\sum_{i=1}^n I\{A_i = a\}}$$

and other variables are defined above. $\hat{V}^{(a)}(t, t)$ is a consistent estimator of $V^{(a)}(t, t)$. For real data with sufficient data points, it is a reliable estimator to construct confidence band.

4 Real Data Analysis

4.1 Binary and Continous Outcomes

We have tried to reproduce the result of this paper on the publicly available NIDA-CTN-0003, NIDA-CTN-0030 and NIDA-CTN-0044 datasets. The code for calculating the estimators is available on their website <https://github.com/BingkaiWang/covariate-adaptive>, whose validity relies on Assumption 1.

Now we verify the two statements of Assumption 1. The first statement states that W_i are i.i.d. samples from an unknown population P , which is always assumed correct. The second statement is the "missing at random" assumption. In all three trials, the outcome of interest is the existence or the proportion of positive urine tests. By assuming missing at random, we assume that conditional on all of the covariates and given the treatment arm, whether a patient misses the

urine test does not rely on the result of the test. The data of CTN-0030 shows that in those who were not performed a urine drug screen on and specified a reason, "participant was unable to provide sample" is mostly chosen. Therefore, as data from observational studies suggest that approximately 10% of episodes of urinary retention might be attributable to the use of concomitant medication[13], the assumption can be seen as suitable.

Our general setting follows the original paper. In all cases, the target of inference is the average treatment effect defined as $E[Y(1)] - E[Y(0)]$. All missing baseline values were imputed by the median for continuous variables and mode for binary or categorical variables. The only estimator in Table 1 that adjusts for missing outcomes is the DR-WLS estimator; all other estimators omit data from the participants with missing outcomes. Negative (positive) estimates are in the direction of clinical benefit (harm). For all estimators presented in Table 1, the 95% CI is constructed using the normal approximation with variance calculated from formula (4). However, since the data used by the paper is unavailable, it is highly likely that the data tidying processes are different.

The following table summarizes our data analyses involving binary and continuous outcomes. The outcome is binary for NIDA-CTN-0003 and continuous for NIDA-CTN-0030 and NIDA-CTN-0044.

	NIDA-CTN-0003	NIDA-CTN-0030	NIDA-CTN-0044
Unadjusted estimator	-0.093 (-0.193,0.008)	0.016 (-0.02,0.053)	-0.096 (-0.152,-0.04)
Adjusted estimator (S)	-0.098 (-0.199,0.002)	0.017 (-0.02,0.053)	-0.092 (-0.148,-0.036)
Adjusted estimator (X)	-0.107 (-0.193,-0.02)	0.015 (-0.018,0.048)	-0.092 (-0.147,-0.037)
DR-WLS estimator (X)	-0.107 (-0.192,-0.021)	0.015 (-0.018,0.048)	-0.098 (-0.155,-0.041)

According to our results, covariate adjustment brings 26%, 17% and 3% variance reduction for NIDA-CTN-0003, NIDA-CTN-0030 and NIDA-CTN-0044, respectively, compared to the unadjusted estimator. This result differs from the original paper on NIDA-CTN-0003, which suggests the variance reduction is 36%. We consider this difference to be due to the unclearness of the outcome selection. In our setting, any of the three types of opioids (methadone, morphine and oxycodone) being tested positive will be considered a sign of a positive urine test.

It is also worth mentioning that when processing NIDA-CTN-0044, the paper asserts that the original strata data cannot be obtained, since the site data, one of the criteria determining the strata, is not available in the dataset; hence, it uses the rest of the criteria to determine the strata. However, the strata itself is actually available in the data. In our setting, we use the actual one, yet get similar results.

4.2 Time-to-Event Outcome

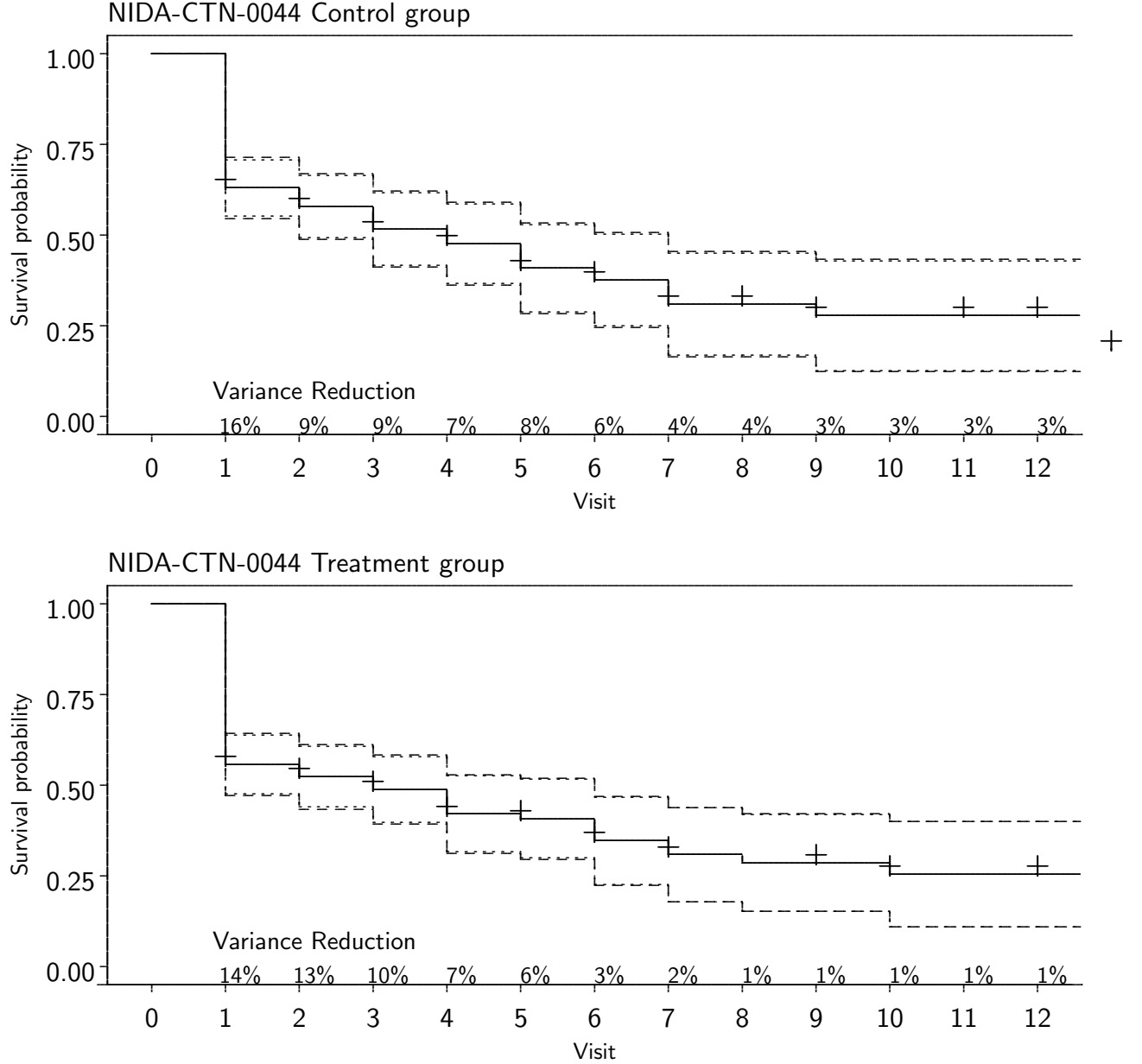
For time-to-event outcomes, assumption 2 is made. Contrary to assumption 1, it assumes censoring completely at random instead of missing at random, and states that there is a positive probability that both the failure time and censoring time occur after the restriction time τ (under each study arm assignment). The latter is often made in survival analysis. As regards the former, it in NIDA-CTN-0044 means the first missing visit is independent of the time to abstinence, defined as the time to the first two consecutive negative urine tests during the study, given the treatment arm.

The following figures (on the next page) present the K-M estimators for time-to-abstinence for the study NIDA-CTN-0044, in the control group and treatment group. The variance of the K-M estimator is estimated in two different ways: one ignored the stratification variable and was the estimated variance returned by the "survfit" function in R; the other used the variance formula proposed by the paper that takes the stratification into account. For each of the two variance estimators, we constructed corresponding point-wise CIs for the K-M estimator.

The solid line is the estimated survival function. Dashed and dotted lines, respectively, represent CIs using the standard method and CIs accounting for randomization strata; the dashed and dotted lines are very similar and almost coincide. The cross signs represent the censoring times. "Variance Reduction" and the associated percentages represent the variance reduction due to accounting for stratified randomization.

While CIs based on different variance estimators are very close to each other, there are variance reductions due to accounting for stratification, which can be translated into sample size reduction needed to achieve the desired power. The variance reduction ranges from 1% to 14% as we consider the survival function at different time points. Among all time points, the first time point (one week after randomization) has the greatest variance reduction.

Owing to the data tidying difference mentioned before, our result is slightly different from that of the original paper (1% to 12%). This result, nevertheless, still proves that the variance reduction exists in this study, and corresponds with the claim of the paper that the variance formulas in this case are asymptotically conservative.



5 Simulation

We in real data analysis compare the adjusted estimators (X) to the unadjusted ones, and show a substantial precision gain. However, in exercise, we may simply adjust the covariates, regardless of if they are the stratification variable, which is to say that the reason for the variance reduction may be the adjustment and the trial design combined. We now in simulations are able to test the impact of randomization methods by comparing the adjusted estimators (X) under covariate-adaptive randomization to that under simple randomization.

5.1 No Missing Data, Binary Outcomes

For binary outcomes, the standardized logistic regression estimator logistic is calculated by: first fitting a working model: $P(Y = 1|A, X) = \text{expit}(\beta_0 + \beta_A A + \beta_X^t X)$, getting the maximum likelihood estimates; then plugging them back into the model to obtain $\hat{\Delta}_{\text{Logistic}} = \frac{1}{N} \sum_{i=1}^N (\text{expit}(\hat{\beta}_0 + \hat{\beta}_A + \hat{\beta}_X^t X_i) - \text{expit}(\hat{\beta}_0 + \hat{\beta}_X^t X_i))$. Now, we

set $N = 300$, the strata $S_i \sim \text{Bernoulli}(\frac{1}{2})$, the covariates $W_{ij} \sim \mathcal{N}(0, 1)$, i.i.d., $X = (W, S)$. With those and other parameters manually fixed, we repeatedly perform simple and stratified randomization ($\pi = 0.2$) and calculate the variances of the estimators using the Bootstrap method[14], with $B = 500$.

The following table shows the variance reduction under 10 tests with parameters randomly changed. Contrary to the predicted results, in a non-negligible proportion of parameter sets, the variance reduction ratios are negative. Hence, the supposed advantages of using covariate-adaptive randomization methods cannot be confirmed under our simulation. However, since the variance and the standard deviation of it is in about the same order of magnitude, it could be the times of simulation, limited by computer resources, are not sufficient to confirm the effect.

	1	2	3	4	5	6	7	8	9	10
Adjusted estimator (X)	0.00200	0.00194	0.00176	0.00181	0.00182	0.00164	0.00188	0.00186	0.00175	0.00196
Estimator under simple randomization	0.00168	0.00180	0.00187	0.00171	0.00173	0.00178	0.00189	0.00196	0.00178	0.00197
Variance reduction	-0.18793	-0.07510	0.05996	-0.05504	-0.05436	0.07695	0.00177	0.05267	0.01846	0.00324

5.2 DR-WLS Estimator, Missing At Random, Binary Outcomes

The DR-WLS estimator is calculated by first fitting the logistic regression working model $P(M = 1|A, X) = \text{expit}(\alpha_0 + \alpha_A A + \alpha_X^t X)$, getting the ML estimates of parameters; then fitting the previous model with weights $1/\text{expit}(\hat{\alpha}_0 + \hat{\alpha}_A A + \hat{\alpha}_X^t X)$ using only the data with $M_i = 1$. With the same parameters and $\alpha_0, \alpha_A, \alpha_X$ manually fixed, we repeatedly perform simple and stratified randomization and calculate the variances of the estimators.

The following table shows the variance reduction under 8 tests with parameters randomly changed. Similar to the previous result, in a non-negligible proportion of parameter sets, the variance reduction ratios are negative. Hence, the supposed advantages of using covariate-adaptive randomization methods with DR-WLS estimators cannot be confirmed under our simulation, either.

	1	2	3	4	5	6	7	8
Adjusted estimator (X)	0.00200	0.00172	0.00201	0.00190	0.00195	0.00251	0.00216	0.00177
Estimator under simple randomization	0.00185	0.00175	0.00188	0.00183	0.00187	0.00201	0.00199	0.00191
Variance reduction	-0.08395	0.01291	-0.06857	-0.03405	-0.04604	-0.24953	-0.08191	0.07355

Considering the results above, it is of little sense to verify the K-M estimator on time-to-event outcomes by our method.

6 Outlook and Discussion

In this report, we reproduce the work of the original paper and perform simulations. Our real data analysis, though slightly different from the original paper, shows the same conclusion that there is apparent variance reduction. Our simulations, unfortunately, are not able to show significant results. More times of simulations may be run.

Theorems in the paper for real-value and time-to-event outcome covers a variety of estimators and provide a heuristic proof sketch, that is proving that under stratified and biased-coin randomization it is asymptotically linear with the same influence function as under simple randomization. M-estimators covers most of useful robust estimator under superpopulation inference framework. For Kaplan-Meier estimator, it is a basic non-parametric estimator of survival function. Future researchers may try to apply the proof approach to other estimators, for example, covariate adjustment K-M estimator in Lu and Tsiatis (2011)[15] and Zhang (2015)[16] or apply the analysis of DR-WLS estimator on augmented estimator, like AIPW-AICW estimator in van der Laan and Robins (2003)[17]. We can analyze real data with those estimators, and analyze real data other than those analyzed in the paper.

References

- [1] EMEA FDA. E9 statistical principles for clinical trials. *US Food and Drug Administration: CDER/CBER. European Medicines Agency: CPMP/ICH/363*, 96, 1998.
- [2] Yunzhi Lin, Ming Zhu, and Zheng Su. The pursuit of balance: an overview of covariate-adaptive randomization techniques in clinical trials. *Contemporary clinical trials*, 45:21–25, 2015.
- [3] Brennan C Kahan and Tim P Morris. Improper analysis of trials randomised using stratified blocks or minimisation. *Statistics in medicine*, 31(4):328–340, 2012.

- [4] Edward L Kaplan and Paul Meier. Nonparametric estimation from incomplete observations. *Journal of the American statistical association*, 53(282):457–481, 1958.
- [5] Marvin Zelen. The randomization and stratification of patients to clinical trials. *Journal of Chronic Diseases*, pages 365–375, 1974.
- [6] Bradley Efron. Forcing a sequential experiment to be balanced. *Biometrika*, 58(3):403–417, 1971.
- [7] Aad W Van der Vaart. *Asymptotic statistics*, volume 3. Cambridge university press, 1998.
- [8] James M Robins, Andrea Rotnitzky, and Lue Ping Zhao. Estimation of regression coefficients when some regressors are not always observed. *Journal of the American statistical Association*, 89(427):846–866, 1994.
- [9] Thomas R Fleming and David P Harrington. *Counting processes and survival analysis*. John Wiley & Sons, 2011.
- [10] Michael R Kosorok. *Introduction to empirical processes and semiparametric inference*. Springer, 2008.
- [11] Federico A Bugni, Ivan A Canay, and Azeem M Shaikh. Inference under covariate-adaptive randomization. *Journal of the American Statistical Association*, 113(524):1784–1796, 2018.
- [12] Federico A Bugni, Ivan A Canay, and Azeem M Shaikh. Inference under covariate-adaptive randomization with multiple treatments. *Quantitative Economics*, 10(4):1747–1785, 2019.
- [13] Verhamme KM, Sturkenboom MC, Stricker BH, and Bosch R. Drug-induced urinary retention: incidence, management and prevention. *Drug Saf.*, 31(5):373–388, 2008.
- [14] B. Efron. Bootstrap Methods: Another Look at the Jackknife. *The Annals of Statistics*, 7(1):1 – 26, 1979.
- [15] Xiaomin Lu and Anastasios A Tsiatis. Semiparametric estimation of treatment effect with time-lagged response in the presence of informative censoring. *Lifetime data analysis*, 17(4):566–593, 2011.
- [16] Min Zhang. Robust methods to improve efficiency and reduce bias in estimating survival curves in randomized clinical trials. *Lifetime data analysis*, 21(1):119–137, 2015.
- [17] Mark J Van der Laan and James M Robins. *Unified methods for censored longitudinal data and causality*, volume 5. Springer, 2003.