Toward Better Practice of Covariate Adjustment in Analyzing Randomized Clinical Trials

Ting Ye[1], Jun Shao[2 3], Yanyao Yi[4], and Qingyuan Zhao[5]

Abstract

In randomized clinical trials, adjustments for baseline covariates at both design and analysis stages are highly encouraged by regulatory agencies. A recent trend is to use a model-assisted approach for covariate adjustment to gain credibility and efficiency while producing asymptotically valid inference even when the model is incorrect. In this article we present three considerations for better practice when model-assisted inference is applied to adjust for covariates under simple or covariate-adaptive randomized trials: (1) guaranteed efficiency gain: a model-assisted method should often gain but never hurt efficiency; (2) wide applicability: a valid procedure should be applicable, and preferably universally applicable, to all commonly used randomization schemes; (3) robust standard error: variance estimation should be robust to model misspecification and heteroscedasticity. To achieve these, we recommend a model-assisted estimator under an analysis of heterogeneous covariance working model that includes all covariates utilized in randomization. Our conclusions are based on an asymptotic theory that provides a clear picture of how covariate-adaptive randomization and regression adjustment alter statistical efficiency. Our theory is more general than the existing ones in terms of studying arbitrary functions of response means (including linear contrasts, ratios, and odds ratios), multiple arms, guaranteed efficiency gain, optimality, and universal applicability.

Keywords: Analysis of covariance; Covariate-adaptive randomization; Efficiency; Heteroscedasticity; Model-assisted; Multiple treatment arms; Treatment-by-covariate interaction.

[1]Department of Biostatistics, University of Washington.
[2]KLATASDS-MOE, School of Statistics, East China Normal University.
[3]Department of Statistics, University of Wisconsin-Madison.
[4]Global Statistical Sciences, Eli Lilly and Company.
[5]Department of Pure Mathematics and Mathematical Statistics, University of Cambridge.
Corresponding to Dr. Jun Shao. Email: shao@stat.wisc.edu.

1 Introduction

Consider a clinical trial with patients randomized into one and only one of multiple treatment arms according to fixed assignment proportions. Each patient has multiple potential responses, one for each treatment, but only one response is observed depending on the assigned treatment. Based on data collected from the trial, we would like to make statistical inference on treatment effects defined as functions of the response means (e.g., linear contrasts, ratios, or odds ratios). These unconditional treatment effects are discussed in a recent Food and Drug Administration (FDA) draft guidance (FDA, 2021).

In clinical trials, we typically observe some baseline covariates for each patient, which are measured prior to treatment assignments and, hence, are not affected by the treatment. As emphasized in regulatory agency guidelines, baseline covariates are encouraged to be utilized in the following two ways. (i) In the design stage, covariate-adaptive randomization can be used to enforce the balance of treatment assignments across levels of discrete baseline prognostic factors, such as institution, disease stage, prior treatment, gender, and age group. "Balance of treatment groups with respect to one or more specific prognostic covariates can enhance the credibility of the results of the trial" (EMA, 2015, European Medicines Agency). (ii) In the analysis stage, baseline covariates can be used to gain efficiency. "Incorporating prognostic baseline factors in the primary statistical analysis of clinical trial data can result in a more efficient use of data to demonstrate and quantify the effects of treatment with minimal impact on bias or the Type I error rate" (FDA, 2021). More specifically, the investigator is advised to "identify those covariates and factors expected to have an important influence on the primary variable(s)" and to specify "how to account for them in the analysis in order to improve precision and to compensate for any lack of balance between groups" (ICH E9, 1998).

For efficiency gain, one may apply a *model-based* approach using a model between potential responses and covariates. However, the validity of a model-based approach requires

a correct model specification, which is a possibly strong assumption. As emphasized in FDA (2021), a method used for covariate adjustment "should provide valid inference under approximately the same minimal statistical assumptions that would be needed for unadjusted estimation in a randomized trial". Consequently, *model-assisted* approaches, which gain efficiency through a working model between responses and covariates and still produce asymptotically valid inference even when the working model is misspecified, have become considerably more popular.

1.1 Considerations in covariate adjustment

For better practice of covariate adjustment via model-assisted approaches, we present the following three considerations.

1. Guaranteed efficiency gain. *The working model should be chosen so that the resulting model-assisted estimator often gains but never loses efficiency when compared to a benchmark estimator that does not adjust for any covariate.*

This consideration is important for model-assisted inference because covariate adjustment based on a misspecified working model does not necessarily lead to efficiency gain over the benchmark. One example is the customary analysis of covariance (ANCOVA) whose working model does not include treatment-by-covariate interaction terms, which we refer to as the *homogeneous* working model (§2.3). These interaction terms are often ignored or even discouraged in practice because of two correct but incomplete perceptions: (i) even if the homogeneous working model is misspecified, ANCOVA still provides valid inference as it is model-assisted; (ii) a model without interaction terms has fewer coefficients to estimate and may have better finite sample properties. However, when the treatment effect is indeed heterogeneous, the ANCOVA estimator using the homogeneous working model may be even less efficient than the benchmark analysis of variance (ANOVA) estimator that uses no model assistance at all (Freedman, 2008a; Lin, 2013). This has led to confusion about how covariate adjustment should be implemented, which can be seen from conflicting rec-

ommendations by regulatory agencies: "The primary model should not include treatment by covariate interactions." (EMA, 2015); "The linear models may include treatment by covariate interaction terms." (FDA, 2021).

Is there a model-assisted method that achieves guaranteed efficiency gain? An affirmative answer is provided in §1.2, followed by theoretical justifications in §3.

2. Wide applicability. *The model-assisted inference procedure should be applicable to all commonly used randomization schemes.*

Covariate-adaptive randomization has been widely used in modern clinical trials to balance treatment assignments across important prognostic factors. According to a review (Ciolino et al., 2019) of nearly 300 clinical trials published in two years, 2009 and 2014, 237 of them used covariate-adaptive randomization. The three most popular covariate-adaptive randomization schemes are the stratified permuted block (Zelen, 1974), the stratified biased coin (Shao et al., 2010; Kuznetsova and Johnson, 2017), and Pocock-Simon's minimization (Taves, 1974; Pocock and Simon, 1975; Han et al., 2009). Unlike simple randomization, covariate-adaptive randomization generates a dependent sequence of treatment assignments. As recognized by regulatory agencies (EMA, 2015; FDA, 2021), conventional inference procedures developed under simple randomization are not necessarily valid under covariate-adaptive randomization. Thus, the second consideration is whether the model-assisted inference procedure is applicable to all commonly used randomization schemes.

3. Robust standard error. *The model-assisted inference should use standard errors robust against model misspecification and heteroscedasticity.*

The use of robust standard error is a crucial step for valid model-assisted inference (FDA, 2021). Although the asymptotic theory for heteroscedasticity-robust standard errors was developed decades ago (Huber, 1967; White, 1980) and has been widely used in econometrics, its usage in clinical trials is unfortunately scarce. Another issue is to take into account covariate centering used in a model-assisted approach.

1.2 Our contributions

In randomized clinical trials, whether and how to adjust for covariates is a long-standing question, which has received revived attention due to influential papers by Yang and Tsiatis (2001), Tsiatis et al. (2008), Freedman (2008a), Lin (2013), and many other publications built upon those. For linear contrast of response means from two treatment arms, theoretical results on efficiency and optimality of various covariate adjustment methods have been developed under simple randomization (Tsiatis et al., 2008; Lin, 2013; Wang et al., 2019; Liu and Yang, 2020, among others) or covariate-adaptive randomization (Shao et al., 2010; Shao and Yu, 2013; Ma et al., 2015; Bugni et al., 2018; Ye, 2018; Ma et al., 2020a,b; Shao, 2021; Wang et al., 2021). Parallel results for linear contrasts of response means from multiple treatment arms under covariate-adaptive randomization are first obtained by Bugni et al. (2019) and Ye et al. (2020) but optimality is not studied.

Given how frequently covariate adjustment is used in practice, it is important to have a guideline regarding the three considerations in §1.1. For this purpose, we establish a comprehensive theory for all aspects of guaranteed efficiency gain, optimality, universal applicability, multiple treatment arms, and variance estimation, which provides insights on how covariate-adaptive randomization and covariate adjustment alter statistical efficiency and facilitates a better practice with clear recommendations for practitioners.

Our theory shows that a *heterogeneous* analysis of covariance working model including all treatment-by-covariate interaction terms and all covariates utilized in covariate-adaptive randomization should be favored over the customary ANCOVA because the former achieves guaranteed efficiency gain, optimality, and wide applicability. To distinguish from the customary ANCOVA that uses a homogeneous working model, we term the analysis of covariance using a heterogeneous working model as ANalysis of HEterogeneous COVAriance (ANHECOVA). Note that ANHECOVA is not a new proposal and has a long history in the literature with a recent resurgence of attention (Cassel et al., 1976; Yang and Tsiatis, 2001;

Tsiatis et al., 2008; Lin, 2013; Wang et al., 2019; Liu and Yang, 2020; Li and Ding, 2020, among others), but our recommendation of ANHECOVA is from a more comprehensive perspective. Specifically, in §3.2-§3.3, we show that under mild and transparent assumptions, the recommended ANHECOVA estimator of the response mean vector is consistent, asymptotically normal, and asymptotically more efficient than the benchmark ANOVA or ANCOVA estimator; in fact, the ANHECOVA estimator is asymptotically optimal within a wide class of linearly-adjusted estimators. A special case of this result for estimating the difference of treatment means under a two-arm trial was obtained independently by Ma et al. (2020b). In §3.1 we offer explanations of why the heterogeneous working model is generally preferable over the homogeneous working model.

Besides guaranteed efficiency gain and wide applicability, our asymptotic theory in §3.2-3.3 shows that the recommended ANHECOVA also enjoys a *universality* property, i.e., the same inference procedure can be universally applied to all commonly used randomization schemes including Pocock-Simon's minimization whose asymptotic property is still not well understood. This is because the asymptotic variance of the ANHECOVA estimator is invariant to the randomization scheme, as long as the randomization scheme satisfies a very mild condition (C2) stated in §2.2. The universality property is desirable for practitioners as they do not need to derive a tailored standard error for each randomization scheme.

The standard heteroscedasticity-robust standard error formulas do not directly apply to model-assisted inference for clinical trials because they do not take into account covariate centering prior to model fitting. In §3.4, we develop a robust standard error formula that can be used with the ANHECOVA estimator.

Finally, our investigation offers new insights on when the customary ANCOVA as a model-assisted inference approach can achieve guaranteed efficiency gain over the benchmark ANOVA. For example, under two-arm trials, Lin (2013) and Ma et al. (2020b) showed that ANCOVA has this desirable property if inference focuses on the linear contrast and the

treatment allocation is balanced. However, our theory shows that this does not extend to trials with more than two arms or inference on nonlinear functions of response means (such as ratios or odds ratios), and is thus a peculiar property for ANCOVA. In addition, ANCOVA is not widely applicable as its asymptotic normality requires an additional condition (C3) on randomization, which is not satisfied by the popular Pocock-Simon's minimization method. Even when ANCOVA is applicable to a particular randomization scheme, it does not have universality because its asymptotic variance varies with the randomization scheme (Bugni et al., 2018).

After introducing the notation, basic assumptions, and working models in §2, we present the methodology and theory in §3. Some numerical results are given in §4. The paper is concluded with recommendations and discussions for clinical trial practice in §5. Technical proofs can be found in the supplementary material.

2 Trial Design and Working Models

2.1 Sample

In a clinical trial with k treatment arms, let $Y^{(t)}$ represent the potential (continuous or discrete) response under treatment t, $t = 1, \ldots, k$, θ be the k-dimensional vector whose tth component is $\theta_t = E(Y^{(t)})$, the unknown population mean of potential response $Y^{(t)}$ under treatment t. We are interested in given functions of θ, such as a linear contrast $\theta_t - \theta_s$, a ratio θ_t/θ_s, or an odds ratio $\{\theta_t/(1-\theta_t)\}/\{\theta_s/(1-\theta_s)\}$ between two treatment arms t and s. We use Z to denote the observed vector of discrete baseline covariates used in covariate-adaptive randomization and X to denote the observed vector of baseline covariates used in model-assisted inference. The vectors Z and X are allowed to share some common entries.

Suppose that a random sample of n patients is obtained from the population under investigation. For the ith patient, let $Y_i^{(1)}, ..., Y_i^{(k)}$, Z_i, and X_i be the realizations of $Y^{(1)}, ..., Y^{(k)}$, Z, and X, respectively. We impose the following mild condition.

(C1) $(Y_i^{(1)}, \ldots, Y_i^{(k)}, Z_i, X_i)$, $i = 1, \ldots, n$, are independent and identically distributed with finite second order moments. The distribution of baseline covariates is not affected by treatment and the covariance matrix $\Sigma_X = \text{var}(X_i)$ is positive definite.

Notice that neither a model between the potential responses and baseline covariates nor a distributional assumption on potential responses is assumed.

2.2 Treatment assignments

Let $\pi_1, \ldots, \pi_k$ be the pre-specified treatment assignment proportions, $0 < \pi_t < 1$, and $\sum_{t=1}^{k} \pi_t = 1$. Let A_i be the k-dimensional treatment indicator vector that equals a_t if patient i receives treatment t, where a_t denotes the k-dimensional vector whose tth component is 1 and other components are 0. For patient i, only one treatment is assigned according to A_i after baseline covariates Z_i and X_i are observed. The observed response is $Y_i = Y_i^{(t)}$ if and only if $A_i = a_t$. Once the treatments are assigned and the responses are recorded, the statistical inference is based on the observed (Y_i, Z_i, X_i, A_i) for $i = 1, ..., n$.

The simple randomization scheme assigns patients to treatments completely at random, under which A_i's are independent of $(Y_i^{(1)}, ..., Y_i^{(k)}, X_i)$'s and are independent and identically distributed with $P(A_i = a_t) = \pi_t$, $t = 1, ..., k$. It does not make use of covariates and, hence, may yield sample sizes that substantially deviate from the target assignment proportions across levels of the prognostic factors.

To improve the credibility of the trial, it is often desirable to enforce the targeted treatment assignment proportions across levels of Z by using covariate-adaptive randomization. As introduced in Section 1, the three most popular covariate-adaptive randomization schemes are the stratified permuted block and stratified biased coin, both of which use all joint levels of Z as strata, and Pocock-Simon's minimization, which aims to enforce treatment assignment proportions across marginal levels of Z.

All these covariate-adaptive randomization schemes, as well as the simple randomization, satisfy the following mild condition (Baldi Antognini and Zagoraiou, 2015).

(C2) The discrete covariate Z used in randomization has finitely many joint levels in $\mathcal{Z}$ and satisfies (i) given $\{Z_i, i = 1, ..., n\}$, $\{A_i, i = 1, ..., n\}$ is conditionally independent of $\{(Y_i^{(1)}, ..., Y_i^{(k)}, X_i), i = 1, ..., n\}$; (ii) as $n \to \infty$, $n_t(z)/n(z) \to \pi_t$ almost surely, where $n(z)$ is the number of patients with $Z = z$ and $n_t(z)$ is the number of patients with $Z = z$ and treatment t, $z \in \mathcal{Z}$, $t = 1, ..., k$.

2.3 Working models

The ANOVA considered as benchmark throughout this paper does not model how the potential responses $Y_i^{(1)}, ..., Y_i^{(k)}$ depend on the baseline covariate vector X_i. It is based on

$$E(Y_i \mid A_i) = \vartheta^T A_i, \tag{1}$$

where ϑ is a k-dimensional unknown vector and c^T denotes the row vector that is the transpose of a column vector c. By Lemma 2 in the supplementary material, ϑ identifies $\theta = (\theta_1, ..., \theta_k)^T$, where $\theta_t = E(Y^{(t)})$ is the mean potential response under treatment t. In the classical exact ANOVA inference, the responses are further assumed to have normal distributions with equal variances. So a common perception is that ANOVA can only be used for continuous responses. As normality is not necessary in the asymptotic theory, the ANOVA and the other approaches introduced next can be used for non-normal or even discrete responses when n is large.

To utilize baseline covariate vector X, the customary ANCOVA is based on the following homogeneous working model,

$$E(Y_i \mid A_i, X_i) = \vartheta^T A_i + \beta^T (X_i - \mu_X), \tag{2}$$

where ϑ and β are unknown vectors having the same dimensions as A and X, respectively, and $\mu_X = E(X_i)$. There is no treatment-by-covariate interaction terms in (2), which is incorrect if patients with different covariates benefit differently from receiving the same treatment, a scenario that often occurs in clinical trials. By Lemma 2 in the supplementary material, $E\{Y_i - \vartheta^T A_i - \beta^T (X_i - \mu_X)\}^2$ is minimized at $(\vartheta, \beta) = (\theta, \beta)$, where $\beta = \sum_{t=1}^{k} \pi_t \beta_t$

and $\beta_t = \Sigma_X^{-1}\text{cov}(X_i, Y_i^{(t)})$. Thus, the ANCOVA estimator with working model (2) is model-assisted (Theorems 1 and 3 in §3). Then, what is the impact of ignoring the treatment-by-covariate interaction effect when it actually exists? The impact is that the ANCOVA estimator may be even less efficient than the benchmark ANOVA estimator, as noted by Freedman (2008a) with some examples.

To better adjust for X, we consider an alternative working model that includes the treatment-by-covariate interactions:

$$E(Y_i \mid A_i, X_i) = \vartheta^T A_i + \sum_{t=1}^{k} \not{\beta}_t^T (X_i - \mu_X) I(A_i = a_t), \tag{3}$$

where $\vartheta, \not{\beta}_1, \ldots, \not{\beta}_k$ are unknown vectors and $I(\cdot)$ is the indicator function. We call model (3) the *heterogeneous* working model because it includes the interaction terms to accommodate the treatment effect heterogeneity across covariates, i.e., patients with different covariate values may benefit differently from treatment. By Lemma 2 in the supplementary material, $E\{Y_i - \vartheta^T A_i - \sum_{t=1}^{k} \not{\beta}_t^T (X_i - \mu_X) I(A_i = a_t)\}^2$ is minimized at $(\vartheta, \not{\beta}_1, ..., \not{\beta}_k) = (\theta, \beta_1, ..., \beta_k)$, where $\beta_t = \Sigma_X^{-1}\text{cov}(X_i, Y_i^{(t)})$, i.e., inference under working model (3) is also model-assisted.

To differentiate the method based on (3) from the ANCOVA based on (2), we refer to the method based on (3) as ANHECOVA.

As a final remark, both working models (2) and (3) use the centered covariate $X - \mu_X$. Centering is crucial to identify θ; the only non-trivial exception is when homogeneous working model (2) is used and linear contrast $\theta_t - \theta_s$ is estimated, as the covariate mean μ_X cancels out. When fitting the working models (2) and (3) with datasets, we can use the least squares with μ_X replaced by $\bar{X}$, the sample mean of all X_i's. In other words, we can center the baseline covariates before fitting the models. Since this step introduces non-negligible variation to the estimation, it affects the asymptotic variance of model-assisted estimator of θ and its estimation for inference. Thus, we cannot assume that the data has been centered in advance and $\mu_X = 0$ without loss of generality (see §3.4).

3 Methodology and Theory

3.1 Estimation

We first describe the estimators of θ under (1)-(3). The ANOVA estimator considered as benchmark is

$$\hat{\theta}_{\text{AN}} = (\bar{Y}_1, ..., \bar{Y}_k)^T, \tag{4}$$

where $\bar{Y}_t$ is the sample mean of the responses Y_i's from patients under treatment t. As $n \to \infty$, $\hat{\theta}_{\text{AN}}$ is consistent and asymptotically normal.

Using the homogeneous working model (2), the ANCOVA estimator of θ is the least squares estimator of the coefficient vector ϑ in the linear model (2) with (A_i, X_i) as regressors. It has the following explicit formula,

$$\hat{\theta}_{\text{ANC}} = \left(\bar{Y}_1 - \hat{\beta}^T(\bar{X}_1 - \bar{X}), ..., \bar{Y}_k - \hat{\beta}^T(\bar{X}_k - \bar{X})\right)^T, \tag{5}$$

where $\bar{X}_t$ is the sample mean of X_i's from patients under treatment t, $\bar{X}$ is the sample mean of all X_i's, and

$$\hat{\beta} = \left\{\sum_{t=1}^{k} \sum_{i:A_i=a_t} (X_i - \bar{X}_t)(X_i - \bar{X}_t)^T\right\}^{-1} \sum_{t=1}^{k} \sum_{i:A_i=a_t} (X_i - \bar{X}_t)Y_i \tag{6}$$

is the least squares estimator of $\not{\beta}$ in (2). It is shown in Theorems 1 and 3 that $\hat{\theta}_{\text{ANC}}$ is consistent and asymptotically normal as $n \to \infty$ regardless of whether working model (2) is correct or not, i.e., ANCOVA is model-assisted.

The term $\hat{\beta}^T(\bar{X}_t - \bar{X})$ in (5) is an adjustment for covariate X applied to the ANOVA estimator $\bar{Y}_t$. However, it may not be the best adjustment in order to reduce the variance. A better choice is to use heterogeneous working model (3). The ANHECOVA estimator of θ is the least squares estimator of ϑ under model (3),

$$\hat{\theta}_{\text{ANHC}} = \left(\bar{Y}_1 - \hat{\beta}_1^T(\bar{X}_1 - \bar{X}), ..., \bar{Y}_k - \hat{\beta}_k^T(\bar{X}_k - \bar{X})\right)^T, \tag{7}$$

where

$$\hat{\beta}_t = \left\{\sum_{i:A_i=a_t} (X_i - \bar{X}_t)(X_i - \bar{X}_t)^T\right\}^{-1} \sum_{i:A_i=a_t} (X_i - \bar{X}_t)Y_i \tag{8}$$

is the least squares estimator of β_t in (3) for each t. It is shown in Theorems 1-3 below that the ANHECOVA estimator $\hat{\theta}_{\text{ANHC}}$ is not only model-assisted, but also asymptotically at least as efficient as $\hat{\theta}_{\text{AN}}$ and $\hat{\theta}_{\text{ANC}}$, regardless of whether model (3) is correct or not.

The following heuristics reveal why the adjustment $\hat{\beta}_t^T(\bar{X}_t - \bar{X})$ in (7) is better than the adjustment $\hat{\beta}^T(\bar{X}_t - \bar{X})$ in (5), and why ANHECOVA often gains but never hurts efficiency even if model (3) is wrong. As the treatment has no effect on X, both $\bar{X}_t$ and $\bar{X}$ estimate the same quantity and, hence, $\hat{\beta}_t^T(\bar{X}_t - \bar{X})$ is an "estimator" of zero. As $n \to \infty$, $\hat{\beta}_t$ converges to $\beta_t = \Sigma_X^{-1}\text{cov}(X, Y^{(t)})$ in probability, regardless of whether (3) is correct or not (Lemma 3 in the supplementary material). Hence, we can "replace" $\hat{\beta}_t^T(\bar{X}_t - \bar{X})$ by $\beta_t^T(\bar{X}_t - \bar{X})$. Under simple randomization,

$$\begin{aligned} \text{var}\{\bar{Y}_t - \beta_t^T(\bar{X}_t - \bar{X})\} &= \text{var}(\bar{Y}_t) + \text{var}\{\beta_t^T(\bar{X}_t - \bar{X})\} - 2\text{cov}\{\bar{Y}_t, \beta_t^T(\bar{X}_t - \bar{X})\} \\ &= \text{var}(\bar{Y}_t) - \text{var}\{\beta_t^T(\bar{X}_t - \bar{X})\}. \end{aligned} \tag{9}$$

Consequently, $\bar{Y}_t - \hat{\beta}_t^T(\bar{X}_t - \bar{X})$ has a smaller asymptotic variance than $\bar{Y}_t$. Note that (9) does not hold with β_t replaced by other quantities. This explains why the adjustment $\hat{\beta}^T(\bar{X}_t - \bar{X})$ in ANCOVA may lose efficiency, as $\hat{\beta}$ in (6) converges to $\pi_1\beta_1 + \cdots + \pi_k\beta_k$.

The variance reduction technique by (9) can be found in the generalized regression (GREG) approach in the survey sampling literature (Cassel et al., 1976; Särndal et al., 2003; Fuller, 2009; Shao and Wang, 2014; Ta et al., 2020). From the theory of GREG, $\hat{\beta}_t$ in (7) may be replaced by any estimator that converges to β_t in probability, without affecting the asymptotic distribution of the GREG estimator. This motivates the following potential improvement to (8), which utilizes the fact that X has the same covariance across treatments and estimates the covariance matrix of X using all patients,

$$\hat{\beta}_t = \frac{n}{n_t}\left\{\sum_{i=1}^{n}(X_i - \bar{X})(X_i - \bar{X})^T\right\}^{-1} \sum_{i:A_i=a_t}(X_i - \bar{X}_t)Y_i, \tag{10}$$

where n_t is the number of units under treatment t. This alternative estimator alleviates the concern of using an unstable inverse in (8) when the sample size is small. In all numerical results in §4, we apply (10) for ANHECOVA.

3.2 Asymptotic theory under simple randomization

We consider asymptotic theory under simple randomization for a general class of estimators of the form

$$\hat{\theta}(\hat{b}_1, ..., \hat{b}_k) = \left(\bar{Y}_1 - \hat{b}_1^T(\bar{X}_1 - \bar{X}) ..., \bar{Y}_k - \hat{b}_k^T(\bar{X}_k - \bar{X})\right)^T, \tag{11}$$

where $\hat{b}_t$'s have the same dimension as X and can either be fixed or depend on the trial data. Note that class (11) contains all estimators we have discussed so far:

$$\hat{\theta}(\hat{b}_1, ..., \hat{b}_k) = \begin{cases} \hat{\theta}_{\text{AN}} & \text{if } \hat{b}_t = 0 \text{ for all } t \\ \hat{\theta}_{\text{ANC}} & \text{if } \hat{b}_t = \hat{\beta} \text{ in (6) for all } t \\ \hat{\theta}_{\text{ANHC}} & \text{if } \hat{b}_t = \hat{\beta}_t \text{ in (8) or (10) for all } t \end{cases} \tag{12}$$

Theorem 1. *Assume (C1) and simple randomization for treatment assignment.*

(i) Assume that $\hat{b}_t \to b_t$ in probability as $n \to \infty$, where b_t is a fixed vector, $t = 1, ..., k$. Then, as $n \to \infty$,

$$\sqrt{n}\left\{\hat{\theta}(\hat{b}_1, ..., \hat{b}_k) - \theta\right\} \to N(0, V_{\text{SR}}(B)) \quad \textit{in distribution,} \tag{13}$$

where

$$V_{\text{SR}}(B) = \text{diag}\{\pi_t^{-1}\text{var}(Y^{(t)} - b_t^T X)\} + \mathscr{B}^T\Sigma_X B + B^T\Sigma_X\mathscr{B} - B^T\Sigma_X B,$$

$\text{diag}(d_t)$ denotes the $k \times k$ diagonal matrix with the tth diagonal element d_t, $\mathscr{B} = (\beta_1, ..., \beta_k)$, the matrix with columns $\beta_1, ..., \beta_k$, and $B = (b_1, ..., b_k)$. In particular, (13) holds for $\hat{\theta}_{\text{AN}}$, $\hat{\theta}_{\text{ANC}}$, and $\hat{\theta}_{\text{ANHC}}$ as described by (12).

(ii) (Optimality of ANHECOVA). $V_{\text{SR}}(B)$ is minimized at $B = \mathscr{B}$ in the sense that $V_{\text{SR}}(B) - V_{\text{SR}}(\mathscr{B})$ is positive semidefinite for all B.

We briefly describe the proof for part (ii) in Theorem 1 and defer other details to the supplementary material. Notice that

$$V_{\text{SR}}(B) - V_{\text{SR}}(\mathscr{B}) = \text{diag}\{\pi_t^{-1}(\beta_t - b_t)^T\Sigma_X(\beta_t - b_t)\} - (\mathscr{B} - B)^T\Sigma_X(\mathscr{B} - B).$$

The positive semidefiniteness of this matrix follows from the following algebraic result with $M = \Sigma_X^{1/2}(\mathscr{B} - B)$.

Lemma 1. *Let M be a matrix whose columns are $m_1, ..., m_k$, and $\pi_1, ..., \pi_k$ be nonnegative constants with $\sum_{t=1}^k \pi_t = 1$. Then $\mathrm{diag}(\pi_t^{-1} m_t^T m_t) - M^T M$ is positive semidefinite.*

We would like to emphasize that Theorem 1(i) holds regardless of whether model (3) is correct or not. Theorem 1(ii) shows that ANHECOVA not only has guaranteed efficiency gain over ANOVA, but is also the most efficient estimator within the class of estimators in (11) as it attains the optimal $V_{\mathrm{SR}}(\mathscr{B})$. Another consequence of Theorem 1(ii) is that adjusting for more covariates in ANHECOVA does not lose and often gains asymptotic efficiency, although adjusting for fewer covariates may have better performance when n is small.

For the important scenario of estimating the linear contrast $\theta_t - \theta_s$ with fixed t and s, the corresponding model-assisted estimator is $c_{ts}^T \hat{\theta}$, where $\hat{\theta}$ is given by (11) and c_{ts} is the k-dimensional vector whose tth component is 1, sth component is -1, and other components are 0. The following corollary provides an explicit comparison of the asymptotic variances of ANOVA, ANCOVA, and ANHECOVA estimators of linear contrasts, showing that the ANHECOVA estimator has strictly smallest asymptotic variance except for some very special cases.

Corollary 1. *Assume (C1) and simple randomization.*

(i) For any t and s, the difference between the asymptotic variances of $\sqrt{n} c_{ts}^T (\hat{\theta}_{\mathrm{AN}} - \theta)$ and $\sqrt{n} c_{ts}^T (\hat{\theta}_{\mathrm{ANHC}} - \theta)$ is

$$\frac{(\pi_s \beta_t + \pi_t \beta_s)^T \Sigma_X (\pi_s \beta_t + \pi_t \beta_s)}{\pi_t \pi_s (\pi_t + \pi_s)} + \frac{(1 - \pi_t - \pi_s)(\beta_t - \beta_s)^T \Sigma_X (\beta_t - \beta_s)}{\pi_t + \pi_s},$$

which is always ≥ 0 with equality holds if and only if

$$\pi_s \beta_t + \pi_t \beta_s = 0 \quad \textit{and} \quad (\beta_t - \beta_s)(1 - \pi_t - \pi_s) = 0. \tag{14}$$

(ii) For any t and s, the difference between the asymptotic variances of $\sqrt{n}c_{ts}^T(\hat{\theta}_{\text{ANC}} - \theta)$ and $\sqrt{n}c_{ts}^T(\hat{\theta}_{\text{ANHC}} - \theta)$ is

$$\frac{(\beta_t - \beta)^T \Sigma_X (\beta_t - \beta)}{\pi_t} + \frac{(\beta_s - \beta)^T \Sigma_X (\beta_s - \beta)}{\pi_s} - (\beta_t - \beta_s)^T \Sigma_X (\beta_t - \beta_s),$$

which is always ≥ 0 with equality holds if and only if

$$\beta = \frac{\pi_s \beta_t + \pi_t \beta_s}{\pi_t + \pi_s} \quad \textit{and} \quad (\beta_t - \beta_s)(1 - \pi_t - \pi_s) = 0. \tag{15}$$

When $k = 2$, i.e., there are only two arms, (14) reduces to $\pi_2\beta_1 + \pi_1\beta_2 = 0$, and (15) reduces to $\beta_1 = \beta_2$ or $\pi_1 = \pi_2 = 1/2$. The same conclusion was also obtained by Lin (2013) under a different framework that only considers the randomness in treatment assignments. Liu and Yang (2020) extended the result in Lin (2013) to stratified simple randomization. This special result has led to a recommendation of ANCOVA over ANHECOVA under two treatment arms with balanced allocation (Wang et al., 2019; Ma et al., 2020b).

However, when there are more than two arms ($k > 2$), (15) holds if and only if $\beta_t = \beta_s = \beta = \sum_{t=1}^k \pi_t \beta_t$, which is a peculiar case. Therefore, $\beta_t = \beta_s$ or balanced allocation is not enough for ANCOVA to be as efficient as ANHECOVA for estimating $\theta_t - \theta_s$. For ANCOVA to have the same asymptotic efficiency as ANHECOVA in estimating $\theta_t - \theta_s$ for *all pairs* of t and s, all β_t's must be the same.

In addition, it follows from Theorem 1 that inference based on ANHECOVA is asymptotically more efficient than that based on ANCOVA when functions of θ other than linear contrasts (such as a ratio or an odds ratio based on two components of θ) are concerned, even in the case of two arms with balanced treatment allocation.

When there are more than two treatment arms, the ANCOVA estimator may hurt efficient compared with the benchmark ANOVA estimator, even under balanced treatment allocation. This is also observed by Freedman (2008a) in some specific examples.

For the comparison of ANHECOVA with benchmark ANOVA, when $k > 2$, (14) holds if and only if $\beta_t = \beta_s = 0$, i.e., X is uncorrelated with $Y^{(t)}$ and $Y^{(s)}$.

3.3 Asymptotic theory under covariate-adaptive randomization

We now consider the estimation of θ under covariate-adaptive randomization as described in §2.2. Specifically, we would like to provide answers to the following two questions: Is there an estimator achieving wide applicability and universality, i.e., the estimator has an asymptotic distribution invariant with respect to all commonly used randomization schemes so that the same inference procedure can be constructed regardless of which randomization scheme is used? Is there an estimator that is asymptotically the most efficient within the class of estimators given by (11) under any covariate-adaptive randomization?

The answers to these two questions are affirmative, as established formally in Theorems 2 and 3, respectively. Importantly, the key to achieve wide applicability and universality as well as guaranteed efficiency gain is using the ANHECOVA estimator $\hat{\theta}_{\text{ANHC}}$ with *all the joint levels of Z included in the covariate X*.

Theorem 2. *(Wide applicability and Universality of ANHECOVA). Assume (C1) and (C2). If heterogeneous model (3) is used with X containing the dummy variables for all the joint levels of Z as a sub-vector, then, regardless of whether working model (3) is correct or not and which randomization scheme is used, as $n \to \infty$,*

$$\sqrt{n}\left(\hat{\theta}_{\text{ANHC}} - \theta\right) \to N\big(0, V_{\text{SR}}(\mathscr{B})\big) \quad \textit{in distribution}, \tag{16}$$

where $V_{\text{SR}}(\mathscr{B}) = \text{diag}\{\pi_t^{-1}\text{var}(Y^{(t)} - \beta_t^T X)\} + \mathscr{B}^T\Sigma_X\mathscr{B}$ and $\mathscr{B} = (\beta_1, ..., \beta_k)$.

Theorem 2 also applies to rerandomization schemes with discrete baseline Z, as rerandomization satisfies (C2) (Li et al., 2018). For two-armed trials under rerandomization, results for model-assisted inference can also be found in Li and Ding (2020, Theorem 3).

Comparing Theorems 1 and 2, we see that the ANHECOVA estimator including all levels of Z has exactly the same asymptotic variance in simple randomization and any covariate-adaptive randomization satisfying (C2), since $V_{\text{SR}}(\mathscr{B})$ is the same as $V_{\text{SR}}(B)$ in (13) with $B = \mathscr{B}$. Therefore, this estimator achieves wide applicability and universality.

As we show next, however, this is not true for ANOVA or ANCOVA using model (2), or for ANHECOVA when Z is not fully included in the working model.

To answer the second question, we need a further condition on the randomization scheme, mainly for estimators not using model (3) or not including all levels of Z in X.

(C3) There exist $k \times k$ matrices $\Omega(z)$, $z \in \mathcal{Z}$, such that, as $n \to \infty$,

$$\sqrt{n}\left(\tfrac{n_1(z)}{n(z)} - \pi_1, \ldots, \tfrac{n_k(z)}{n(z)} - \pi_k,\, z \in \mathcal{Z}\right)^T \mid Z_1, \ldots, Z_n \to N(0, D) \quad \text{in distribution,}$$

where D is a block diagonal matrix whose blocks are matrices $\Omega(z)/P(Z_i = z), z \in \mathcal{Z}$.

Condition (C3) weakens Assumption 4.1(c) of Bugni et al. (2019) in which $\Omega(z)$ takes a more special form. For simple randomization, $\Omega(z) = \text{diag}(\pi_t) - \pi\pi^T$ for all z, where $\pi = (\pi_1, \ldots, \pi_k)^T$. For stratified permuted block randomization and stratified biased coin randomization, $\Omega(z) = 0$ for all z. Note that Pocock-Simon's minimization scheme does not satisfy (C3) because the treatment assignments are correlated across strata, although some recent theoretical result has been obtained (Hu and Zhang, 2020). Thus, the following result does not apply to Pocock-Simon's minimization. However, our Theorem 2 applies to minimization, as (C3) is not needed in Theorem 2.

The next theorem establishes the asymptotic distributions of estimators in class (11) under covariate-adaptive randomization, based on which we show the optimality of the ANHECOVA estimator.

Theorem 3. *Assume (C1), (C2), and (C3). Consider class (11) of estimators and, without loss of generality, we assume that all levels of Z are included in X, as the components of $\hat{b}_t$'s in (11) corresponding to levels of Z not in X may be set to 0.*

(i) For $\hat{\theta}(\hat{b}_1, ..., \hat{b}_k)$ defined in (11) with $\hat{b}_t \to b_t$ in probability as $n \to \infty$, $t = 1, ..., k$,

$$\sqrt{n}\left\{\hat{\theta}(\hat{b}_1, ..., \hat{b}_k) - \theta\right\} \to N\big(0, V(B)\big) \quad \textit{in distribution,} \tag{17}$$

where

$$V(B) = V_{\mathrm{SR}}(B) - E\left[R(B)\{\Omega_{\mathrm{SR}} - \Omega(Z_i)\}R(B)\right], \tag{18}$$

$V_{\mathrm{SR}}(B)$ *is given in (13),* $B = (b_1, ..., b_k)$, $\Omega_{\mathrm{SR}} = \mathrm{diag}(\pi_t) - \pi\pi^T$, *and* $R(B) = \mathrm{diag}\big(\pi_t^{-1} E\{Y_i^{(t)} - \theta_t - b_t^T(X_i - \mu_X) \mid Z_i\}\big)$. *Furthermore,* $R(\mathscr{B}) = 0$ *and, hence,* $V(\mathscr{B}) = V_{\mathrm{SR}}(\mathscr{B})$, *where* $\mathscr{B} = (\beta_1, ..., \beta_k)$.

(ii) (Optimality of ANHECOVA). $V(B)$ *is minimized at* $B = \mathscr{B}$ *in the sense that* $V(B) - V(\mathscr{B})$ *is positive semidefinite for all* B.

The main technical challenge in the proofs of Theorem 2 and Theorem 3 is that the treatment assignments $A_1, \ldots, A_n$ are not independent due to covariate-adaptive randomization, so we cannot directly apply the classical Linderberg central limit theorem. Instead, we decompose $\hat{\theta}(\hat{b}_1, ..., \hat{b}_k) - \theta$ into four terms and then apply a conditional version of the Linderberg central limit theorem to handle the dependence. The details can be found in the supplementary material.

A number of conclusions can be made from Theorem 3.

1. With Theorem 2 answering the first question in the beginning of §3.3, i.e., $\hat{\theta}_{\mathrm{ANHC}}$ with all joint levels of Z included in model (3) achieves wide applicability and universality, the second question is answered by Theorem 3(ii) showing that $\hat{\theta}_{\mathrm{ANHC}}$ is asymptotically the most efficient estimator compared with all estimators in class (11); in particular, $\hat{\theta}_{\mathrm{ANHC}}$ attains guaranteed efficiency gain under any covariate-adaptive randomization satisfying (C2). Our optimality result in Theorem 3(ii) is about the joint estimation of the vector θ, which is substantially more general than the existing one-dimensional optimality results about linear contrasts. Furthermore, our conclusion made in §3.2, i.e., ANHECOVA is asymptotically superior over ANCOVA except for the particular case of estimating a linear contrast for two arms with balanced treatment allocation, holds for all commonly used covariate-adaptive randomization schemes.

2. A price paid for not using model (3) or not including all levels of Z in (3) is that the asymptotic validity of the resulting estimator requires condition (C3), which is not

needed in Theorem 2. Furthermore, the resulting estimator not only is less efficient according to the previous conclusion, but also has a more complicated asymptotic covariance matrix depending on the randomization schemes (universality is not satisfied), which requires extra handling in variance estimation for inference; see, for example, Shao et al. (2010), Bugni et al. (2018), and Ma et al. (2020a).

3. Under covariate-adaptive randomization satisfying (C2)-(C3), it is still true that the ANCOVA estimator using model (2) may be asymptotically more efficient or less efficient than the benchmark ANOVA estimator.

4. From (18), the asymptotic covariance matrix $V(B)$ is invariant with respect to randomization scheme if $R(B)$ in (18) is 0, which is the case when $B = \mathscr{B}$, i.e., $\hat{\theta}_{\text{ANHC}}$ is used with all levels of Z included in X. If $R(B)$ is not 0, such as the case of ANOVA, ANCOVA, or ANHECOVA not adjusting for all joint levels of Z, then $V(B)$ depends on randomization scheme and, the smaller the $\Omega(z)$, the more efficient the estimator is. Thus, the stratified permuted block or biased coin with $\Omega(z) = 0$ for all z is preferred in this regard.

5. The roles played by design and modeling can be understood through

$$V(B) - V_{\text{SR}}(0) = \{V_{\text{SR}}(B) - V_{\text{SR}}(0)\} - E\left[R(B)\{\Omega_{\text{SR}} - \Omega(Z_i)\}R(B)\}\right],$$

where $V_{\text{SR}}(0)$ is the asymptotic variance of ANOVA estimator under simple randomization. As we vary the randomization scheme and the working model, the change in the asymptotic variance is determined by two terms. The first term $\{V_{\text{SR}}(B) - V_{\text{SR}}(0)\}$ arises from using a working model; the second term $E\left[R(B)\{\Omega_{\text{SR}} - \Omega(Z_i)\}R(B)\right]$ is the reduction due to using a covariate-adaptive randomization scheme, which also depends on the working model being used via $R(B)$. Therefore, it is interesting to note that although the primary reason of using covariate-adaptive randomization is to achieve balance of treatment groups across prognostic factors, it also improves statistical efficiency.

Theorem 3 together with a further derivation leads to the following result.

Corollary 2 (Duality between design and analysis)**.** *Assume (C1)-(C3) and that X only includes the dummy variables for all joint levels of Z. Then, for any B in (17), $V(B) = V_{\mathrm{SR}}(\mathscr{B}) + E\{R(B)\Omega(Z_i)R(B)\}$.*

A direct consequence from Corollary 2 is that, if $\Omega(z) = 0$ for all z (e.g., stratified permuted block or biased coin randomization is used) and X *only* includes all joint levels of Z, then all estimators in class (11), including the benchmark ANOVA estimator, have the same asymptotic efficiency as the ANHECOVA estimator under any randomization. This shows the duality between design and analysis, i.e., modeling with all joint levels of Z is equivalent to designing with Z.

3.4 Robust standard error

For model-assisted inference on a function of θ based on Theorems 1-3, a crucial step is to construct a consistent estimator of asymptotic variance. The customary linear model-based variance estimation assuming homoscedasticity can be inconsistent, as criticized by Freedman (2008a) and FDA (2021). Therefore, it is important that we use variance estimators that are consistent regardless of whether the working model is correct or not and whether heteroscedasticity is present or not.

Consider the ANHECOVA estimator $\hat{\theta}_{\mathrm{ANHC}}$ in (7) using either (8) or (10), where covariate X includes all dummy variables for Z that is used in the randomization. There exist formulas for heteroscedasticity-robust standard error (such as those provided in the `sandwich` package in `R`). However, those formulas cannot be directly applied here, because they do not account for the additional variation introduced by centering the covariate X as required by the identification of θ. In fact, the term $\mathscr{B}^T\Sigma_X\mathscr{B}$ in the asymptotic variance $V_{\mathrm{SR}}(\mathscr{B})$ in Theorem 2 arises from centering X.

Instead, we should use the robust variance estimator based on $V_{\rm SR}(\mathscr{B})$, as described next. Let $\hat{\Sigma}_X$ be the sample covariance matrix of X_i based on the entire sample and $S_t^2(\hat{\beta}_t)$ be the sample variance of $(Y_i - \hat{\beta}_t^T X_i)$'s based on the patients in treatment arm t. Then $V_{\rm SR}(\mathscr{B})$ in (16) can be estimated by

$$\hat{V} = \mathrm{diag}\{\pi_t^{-1} S_t^2(\hat{\beta}_t)\} + \hat{\mathscr{B}}^T \hat{\Sigma}_X \hat{\mathscr{B}}, \tag{19}$$

where $\hat{\mathscr{B}}$ is $\mathscr{B}$ with β_t replaced by $\hat{\beta}_t$. This variance estimator is consistent as $n \to \infty$ regardless of whether the heterogeneous working model (3) or homoscedasticity holds or not, and regardless of which randomization scheme is used. When the sample size n is not very large, a correction suggested by MacKinnon and White (1985) may be applied. For estimation or inference about a differentiable function of θ, a robust variance estimator can be obtained based on (19) and the delta method.

In many applications the primary analysis is about treatment effects in terms of the linear contrast $\theta_t - \theta_s = c_{ts}^T \theta$ for one or several pairs of (t, s). For large n, an asymptotic level $(1-\alpha)$ confidence interval of $\theta_t - \theta_s$ is

$$\left(c_{ts}^T \hat{\theta}_{\rm ANHC} - z_{\alpha/2} \mathrm{SE}_{ts},\ c_{ts}^T \hat{\theta}_{\rm ANHC} + z_{\alpha/2} \mathrm{SE}_{ts} \right),$$

where $\mathrm{SE}_{ts}^2 = \pi_t^{-1} S_t^2(\hat{\beta}_t) + \pi_s^{-1} S_s^2(\hat{\beta}_s) + (\hat{\beta}_t - \hat{\beta}_s)^T \hat{\Sigma}_X (\hat{\beta}_t - \hat{\beta}_s)$ and z_α is the $(1-\alpha)$ quantile of the standard normal distribution. The same form of confidence interval can be used for any linear contrast $c^T\theta$ (the sum of components of c is 0) with $c_{ts}^T \hat{\theta}_{\rm ANHC}$ and SE_{ts}^2 replaced by $c^T \hat{\theta}_{\rm ANHC}$ and $\mathrm{SE}_c^2 = c^T \hat{V} c$, respectively. Let $\mathscr{C}$ be the collection of all linear contrasts with dimension k. An asymptotic level $(1-\alpha)$ simultaneous confidence band of $c^T\theta$, $c \in \mathscr{C}$, can be obtained by Scheffé's method,

$$\left(c^T \hat{\theta}_{\rm ANHC} - \chi_{\alpha,k-1}\, \mathrm{SE}_c,\ c^T \hat{\theta}_{\rm ANHC} + \chi_{\alpha,k-1}\, \mathrm{SE}_c \right), \quad c \in \mathscr{C},$$

where $\chi_{\alpha,k-1}$ is the square root of the $(1-\alpha)$ quantile of the chi-square distribution with $(k-1)$ degrees of freedom. Correspondingly, to test the hypothesis $H_0 : \theta_1 = \cdots = \theta_k$, an asymptotic level α chi-square test rejects H_0 if and only if

$$\hat{\theta}_{\rm ANHC}^T C^T (C \hat{V} C^T)^{-1} C \hat{\theta}_{\rm ANHC} > \chi_{\alpha,k-1}^2,$$

where C is the $(k-1) \times k$ matrix whose tth row is c_{tk}^T, $t = 1, ..., k-1$.

Inference procedures based on the ANOVA or ANCOVA estimator can be similarly obtained using Theorems 1 and 3. However, as they do not achieve universality, a tailored derivation is needed for each covariate-adaptive randomization scheme. For example, under the stratified permuted block or biased coin randomization, the ANOVA or ANCOVA estimator is asymptotically more efficient than the same estimator under simple randomization; thus, using variance estimators valid only under simple randomization may lead to unduly conservative inference (FDA, 2021). To eliminate the conservativeness, modifications depending on covariate-adaptive randomization schemes have to be made (Shao et al., 2010; Bugni et al., 2018). For Pocock-Simon's minimization, however, how to derive the tailored variance estimators for the ANOVA and ANCOVA estimators is not yet known as the asymptotic properties of the minimization scheme is still not well established. This is why we recommend ANHECOVA over the other model-assisted estimators for the practice.

4 Empirical Results

4.1 Simulation results

To examine the finite-sample properties of the model-assisted procedures, we perform a simulation study based on the placebo arm of 481 patients in a clinical trial for rheumatoid arthritis. We obtain from the trial a 2-dimensional continuous baseline covariate (U, W) and a continuous response variable $Y^{(1)}$, where U is the baseline disease activity score for rheumatoid arthritis, W is patient's tender joint count at 68 joints, and $Y^{(1)}$ is the change of disease activity score from baseline. The mean vector and covariance matrix of $(Y^{(1)}, U, W)$ based on 481 patients are given in the supplementary material. Note that we do not know the true relationship between $Y^{(1)}$ and (U, W). In fact, a linear model fit between $Y^{(1)}$ and (U, W) based on 481 patients results in multiple and adjusted R-squares ≤ 0.05. Thus,

working models (2) and (3) are likely to be misspecified in our simulation.

We consider two simulation settings that are different in how the potential responses $Y^{(2)}$ and $Y^{(3)}$ of the other two treatment arms are generated, and how the treatment assignment is randomized. Our first simulation compares the standard deviations of the ANOVA, ANCOVA, and ANHECOVA estimators of $\theta_2 - \theta_1$, with $X = U$ for ANCOVA and ANHECOVA. The two additional potential responses are generated according to

$$\begin{aligned} Y^{(2)} &= Y^{(1)} + \zeta(U - \mu_U) \\ Y^{(3)} &= Y^{(2)} \end{aligned} \tag{20}$$

($\theta_1 = \theta_2 = \theta_3$, i.e., no average treatment effect). The sample size is 481 (all data points are sampled). Treatments are assigned by simple randomization according to three different allocation proportions: 1:2:2, 1:1:1, and 2:1:1. Thus, the only randomness in the first simulation is from treatment assignments. Since $\beta_t = \Sigma_X^{-1}\text{cov}(X_i, Y_i^{(t)})$, $\beta_2 = \beta_3 = \beta_1 + \zeta$. The value of β_1 is -0.255 and the value of ζ represents the magnitude of treatment-by-covariate interaction. But ζ does not affect the average treatment effect as it is the coefficient in front of centered $U - \mu_U$. Although we only consider the estimation of $\theta_2 - \theta_1$, data from the third arm is still used by ANCOVA and ANHECOVA.

Based on 10,000 simulations, all three estimators have negligible biases and their standard deviations are plotted in Figure 1 for values of ζ between -1 and 1. The simulation result shows that, as predicted by our theory, ANHECOVA is generally more efficient than the other two estimators, except when ζ is nearly 0 where ANCOVA is comparable to ANHECOVA. Furthermore, the simulation with allocation 1:2:2 (left panel in Figure 1) shows very clearly that there is no definite ordering of the variances of ANCOVA and ANOVA. Corollary 1 suggests that a balanced allocation does not guarantee the superiority of ANCOVA over ANOVA when there are multiple arms (in contrast with the case of two arms), which can be seen from the simulations with allocation 1:1:1 (middle panel in Figure 1).

The second simulation setting is intended to examine the performance of estimators,

standard errors, and the proposed 95% asymptotic confidence intervals for linear contrasts $\theta_2 - \theta_1$ and $\theta_3 - \theta_1$ and ratios θ_2/θ_1 and θ_3/θ_1, under three randomizations schemes, simple randomization, stratified permuted block, and Pocock-Simon's minimization, with allocation 1:1:1 or 1:2:2. For each simulation, a random sample of size 400 is drawn from the 481 subjects' $(Y^{(1)}, U, W)$ with replacement, and $Y^{(2)}$ and $Y^{(3)}$ are generated according to

$$\begin{aligned} Y^{(2)} &= -1.3 + Y^{(1)} - 0.5(U - \mu_U) - 0.01(U^2 - \mu_{U^2}) + 0.3(W - \mu_W) \\ Y^{(3)} &= -1 + Y^{(1)} - 0.1(U - \mu_U) - 0.01(U^2 - \mu_{U^2}) - 0.1(W - \mu_W) \end{aligned} \tag{21}$$

($\theta_1 = -1.031$, $\theta_2 = -2.331$, and $\theta_3 = -2.031$). The magnitude of treatment-by-covariate interaction is represented by the differences of β_t-values, where for $X = (U, W)^T$, $\beta_1 = (-0.240, -0.001)^T$, $\beta_2 = (-0.853, 0.298)^T$, and $\beta_3 = (-0.453, -0.102)^T$. Note that a quadratic term $U^2 - \mu_{U^2}$ appears in the data generating process (21) but is not adjusted by ANCOVA or ANHECOVA. Thus, the models for $Y^{(2)} - Y^{(1)}$ and $Y^{(3)} - Y^{(1)}$ are also misspecified, in addition to the likely event that the model for $Y^{(1)}$ is misspecified.

The covariate Z used in randomization is composed of a three-level discretized W (with proportions 0.24, 0.22, and 0.54) and a two-level discretized U (with proportions 0.77 and 0.23). These Z-categories are created according to the disease activity encoded by covariates U and W. For stratified permuted block randomization, block sizes are 6 and 10 for treatment allocations 1:1:1 and 1:2:2, respectively. For minimization, we follow the procedure in Pocock and Simon (1975), which assigns a patient with probability 0.8 to the preferred arm minimizing the sum of assignment balance scores over marginal levels of Z.

For ANCOVA and ANHECOVA, we consider two working models with different choices of X. One model includes the dummy variables for Z but not (U, W), motivated by the fact that Z is a discretization of (U, W). The other model includes not only the dummy variables for Z, but also U and W. The simulation results with $n = 400$ based on 10,000 simulations are shown in Tables 1 and 2 for linear contrasts and ratios, respectively. When ratio is considered, we apply the delta method to construct standard errors. Simulation

results with $n = 200$ are in the supplementary material.

Note that in the second simulation when covariate-adaptive randomization is used, for ANOVA or ANCOVA, we employ the standard error derived under simple randomization based on Theorem 1. According to our theory, it is expected that the standard errors and the related confidence intervals based on ANOVA and ANCOVA are conservative; the simulation can show how serious the conservativeness is.

The following is a summary of simulation results in Tables 1 and 2.

1. All estimators have negligible bias compared to their standard deviation.
2. ANHECOVA has the smallest standard deviation in all scenarios of our simulation. This is consistent with our asymptotic theory.
3. There is no unambiguous ordering of the standard deviations of ANCOVA and ANOVA. In particular, ANCOVA is better in estimating $\theta_2 - \theta_1$ but worse in estimating $\theta_3 - \theta_1$. However, for allocation 1:2:2 and estimating of ratios, ANCOVA is nearly the same or worse than ANOVA.
4. For ANHECOVA, including additional covariates U and W in the working model results in a smaller standard deviation, indicating that U and W carry more information than their discretized values. But this is not always the case for ANCOVA.
5. From Tables 1 and 2, the performances of ANHECOVA are nearly the same under simple randomization, stratified permuted block, and Pocock-Simon's minimization. This supports the universality results in our asymptotic theory.
6. Under simple randomization, the robust standard errors for all model-assisted estimators are very close to their actual standard deviations, and confidence intervals have nominal coverage in all settings. However, although this is still true for ANHECOVA under stratified permuted block and Pocock-Simon's minimization, it is not the case for ANOVA and ANCOVA, i.e., standard errors valid under simple randomization appear to overestimate the actual standard deviations, so the confidence intervals are conservative. This observation reflects the universality property of ANHECOVA.

4.2 A real data example

We further illustrate the different model-assisted procedures using a real data example. Chong et al. (2016) conducted a randomized experiment to evaluate the impact of low dietary iron intake on human capital attainment. They recruited students of age 11 to 19 in a rural area of Cajamarca, Peru, where many adolescents suffer from iron deficiency. The goal of this randomized trial is to quantify the causal effect of reduced adolescent anemia on school attainment. By using students' school grade as covariate Z with five levels, a stratified permuted block randomization with 1:1:1 allocation was applied to assign 219 students to one of the following three promotional videos:

Video 1: A popular soccer player is encouraging iron supplements to maximize energy;

Video 2: A physician is encouraging iron supplements for overall health;

Video 3: A dentist encouraging oral hygiene without mentioning iron at all.

Chong et al. (2016) investigated whether showing different promotional videos to the students can improve their academic performance through increased iron intake. Video 3 is treated as a "placebo". After the treatment assignments, four students were excluded from the analysis for various reasons, which we also ignore in our analysis. The dataset is available at `https://www.openicpsr.org/openicpsr/project/113624/version/V1/view`.

Chong et al. (2016) used various outcomes in their analysis; here we focus on one of their primary outcomes—the academic achievement—as an example. In this trial, the academic achievement is measured by a standardized average of the student's grades in math, foreign language, social sciences, science, and communications in a semester. For the model-assisted approaches, we use the baseline anemia status as the covariate in working models (2) and (3), which is believed to moderate the treatment effect (Chong et al., 2016).

Table 3 reports the analysis results by using different model-assisted procedures. Like in our simulation studies, the standard errors for ANOVA and ANCOVA are computed using estimator based on Theorem 1 for simple randomization, even though the randomization

scheme here is covariate-adaptive. All the model-assisted estimators find similar effect sizes for the two contrasts (physician versus placebo, soccer star versus placebo), and the two ANHECOVA estimators have slightly smaller standard errors. Including baseline anemia status in the working model is useful to reduce the standard error. Compared to the placebo, the promotional video by the soccer player does not seem to have a positive effect on the academic achievement. In contrast, the video of the physician promoting iron supplements appears to have a positive effect. The difference between ANHECOVA and ANOVA or ANCOVA, and between including and not including anemia can be seen from the magnitude of the corresponding p-values.

5 Recommendation and Discussion

To improve its credibility and efficiency, we believe a clinical trial analysis can benefit from the considerations outlined in §1.1 and discussed throughout §2-3.

Our theoretical investigation shows that the ANHECOVA with all joint levels of Z included in heterogeneous working model (3), coupled with the robust variance estimator given by (19), achieves guaranteed efficiency gain over benchmark ANOVA, asymptotic optimality among a large class of estimators, wide applicability and universality to a wide range of covariate-adaptive randomization schemes. Our theory is for the joint asymptotic distribution in estimating θ (the vector of mean responses), which can be readily used for inference about linear or nonlinear functions of θ. Thus, we believe it deserves wider usage in the clinical trial practice. In addition to all joint levels of Z, other baseline covariates highly associated with the responses can also be included in the ANHECOVA working model, following the guidance of FDA (2021). Our theory shows that using ANOVA, ANCOVA with model (2), or ANHECOVA that does not adjust for all joint levels of Z, suffers from invalidity, inefficiency, or non-universality in the sense that the asymptotic distribution of the estimator depends on a particular randomization scheme.

For discrete responses, although ANHECOVA and our theory can still be applied, it is more common to perform covariate adjustment with nonlinear working models such as the g-computation. Under simple randomization, extensive developments for g-computation can be found in Freedman (2008b), Moore and van der Laan (2009), Rosenblum and van der Laan (2010), Steingrimsson et al. (2017), and Guo and Basse (2021). Recently, Wang et al. (2021) obtained some general results under stratified biased coin and permuted block randomization, including robust inference on a linear contrast of θ using logistic regression as a working model. However, estimators from these methods are not guaranteed to gain efficiency over the unadjusted estimator. We plan to develop covariate adjustment methods with nonlinear working models that achieve all three considerations as a future work.

Multiple treatment arms, which usually include a placebo, different doses (or regimens) of a new treatment, and/or active controls, are common in clinical trials (Juszczak et al., 2019) and are prevalent in some therapeutic areas such as immunology (Yates et al., 2021). In some applications, the primary analysis may focus on comparing just two treatments, even though the trial contains more than two treatment arms. A simple strategy is to ignore the data from other arms and apply inference procedures to the two arms of interest. For ANOVA, this is equivalent to using all the arms, since ANOVA does not borrow strength from other arms through using covariates. However, using data from all arms is recommended for ANHECOVA, because it utilizes covariate data from arms other than the two arms of interest to gain efficiency (implied by Theorem 2). Regarding ANCOVA, there is no definite order of efficiency for using the whole dataset or data from two given arms, since using more covariate data in ANCOVA may increase or decrease efficiency.

As a final cautionary note, standard software does not produce asymptotically valid standard errors for model-assisted inference. We implement an R package called RobinCar to compute the model-assisted estimators and their robust standard errors, which is available at `https://github.com/tye27/RobinCar`.

Supplementary Material

The supplementary material contains additional simulation results and all technical proofs.

Acknowledgements

The authors would like to thank the Editor, Associate Editor and two anonymous referees for helpful comments and suggestions.

Funding

Jun Shao's research was partially supported by research grants from the National Natural Science Foundation of China and the U.S. National Science Foundation. Qingyuan Zhao was supported by a research grant from the Isaact Newton Trust.

References

Baldi Antognini, A. and Zagoraiou, M. (2015). On the almost sure convergence of adaptive allocation procedures. *Bernoulli Journal*, 21(2):881–908.

Bugni, F. A., Canay, I. A., and Shaikh, A. M. (2018). Inference under covariate-adaptive randomization. *Journal of the American Statistical Association*, 113(524):1784–1796.

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.

Cassel, C. M., Särndal, C. E., and Wretman, J. H. (1976). Some results on generalized difference estimation and generalized regression estimation for finite populations. *Biometrika*, 63(3):615–620.

Chong, A., Cohen, I., Field, E., Nakasone, E., and Torero, M. (2016). Iron deficiency and schooling attainment in peru. *American Economic Journal: Applied Economics*, 8(4):222–255.

Ciolino, J. D., Palac, H. L., Yang, A., Vaca, M., and Belli, H. M. (2019). Ideal vs. real: a systematic review on handling covariates in randomized controlled trials. *BMC Medical Research Methodology*, 19(1):136.

EMA (2015). Guideline on adjustment for baseline covariates in clinical trials. Committee for Medicinal Products for Human Use, European Medicines Agency (EMA).

FDA (2021). Adjusting for covariates in randomized clinical trials for drugs and biological products. Draft Guidance for Industry. Center for Drug Evaluation and Research and Center for Biologics Evaluation and Research, Food and Drug Administration (FDA), U.S. Department of Health and Human Services. May 2021.

Freedman, D. A. (2008a). On regression adjustments in experiments with several treatments. *Annals of Applied Statistics*, 2(1):176–196.

Freedman, D. A. (2008b). Randomization does not justify logistic regression. *Statistical Science*, 23(2):237–249.

Fuller, W. A. (2009). *Sampling Statistics*. Wiley, New York.

Guo, K. and Basse, G. (2021). The generalized oaxaca-blinder estimator. *Journal of the American Statistical Association*, 116:1–13.

Han, B., Enas, N. H., and McEntegart, D. (2009). Randomization by minimization for unbalanced treatment allocation. *Statistics in Medicine*, 28(27):3329–3346.

Hu, F. and Zhang, L.-X. (2020). On the theory of covariate-adaptive designs. *arXiv preprint arXiv:2004.02994*.

Huber, P. J. (1967). The behavior of maximum likelihood estimates under nonstandard conditions. In *Proceedings of the Fifth Berkeley symposium on mathematical statistics and probability*, volume 1, pages 221–233. University of California Press.

ICH E9 (1998). Statistical principles for clinical trials E9. International Council for Harmonisation (ICH).

Juszczak, E., Altman, D. G., Hopewell, S., and Schulz, K. (2019). Reporting of multi-arm parallel-group randomized trials: Extension of the consort 2010 statement. *JAMA*, 321(16):1610–1620.

Kuznetsova, O. M. and Johnson, V. P. (2017). Approaches to expanding the two-arm biased coin randomization to unequal allocation while preserving the unconditional allocation ratio. *Statistics in Medicine*, 36(16):2483–2498.

Li, X. and Ding, P. (2020). Rerandomization and regression adjustment. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 82(1):241–268.

Li, X., Ding, P., and Rubin, D. B. (2018). Asymptotic theory of rerandomization in treatment–control experiments. *Proceedings of the National Academy of Sciences*, 115(37):9157–9162.

Lin, W. (2013). Agnostic notes on regression adjustments to experimental data: Reexamining freedman's critique. *Annals of Applied Statistics*, 7(1):295–318.

Liu, H. and Yang, Y. (2020). Regression-adjusted average treatment effect estimates in stratified randomized experiments. *Biometrika*, 107(4):935–948.

Ma, W., Hu, F., and Zhang, L. (2015). Testing hypotheses of covariate-adaptive randomized clinical trials. *Journal of the American Statistical Association*, 110(510):669–680.

Ma, W., Qin, Y., Li, Y., and Hu, F. (2020a). Statistical inference for covariate-adaptive randomization procedures. *Journal of the American Statistical Association*, 115(531):1488–1497.

Ma, W., Tu, F., and Liu, H. (2020b). Regression analysis for covariate-adaptive randomization: A robust and efficient inference perspective. *arXiv preprint arXiv:2009.02287*.

MacKinnon, J. G. and White, H. (1985). Some heteroskedasticity-consistent covariance matrix estimators with improved finite sample properties. *Journal of econometrics*, 29(3):305–325.

Moore, K. L. and van der Laan, M. J. (2009). Covariate adjustment in randomized trials with binary outcomes: targeted maximum likelihood estimation. *Statistics in Medicine*, 28(1):39–64.

Pocock, S. J. and Simon, R. (1975). Sequential treatment assignment with balancing for prognostic factors in the controlled clinical trial. *Biometrics*, 31(1):103–115.

Rosenblum, M. and van der Laan, M. J. (2010). Simple, efficient estimators of treatment effects in randomized trials using generalized linear models to leverage baseline variables. *The International Journal of Biostatistics*, 6(1):13.

Särndal, C.-E., Swensson, B., and Wretman, J. (2003). *Model Assisted Survey Sampling*. Springer Science & Business Media.

Shao, J. (2021). Inference for covariate-adaptive randomization: Aspects of methodology and theory (with discussions). *Statistical Theory and Related Fields*, 5(3):172–186.

Shao, J. and Wang, S. (2014). Efficiency of model-assisted regression estimators in sample surveys. *Statistica Sinica*, 24(1):395–414.

Shao, J. and Yu, X. (2013). Validity of tests under covariate-adaptive biased coin randomization and generalized linear models. *Biometrics*, 69(4):960–969.

Shao, J., Yu, X., and Zhong, B. (2010). A theory for testing hypotheses under covariate-adaptive randomization. *Biometrika*, 97(2):347–360.

Steingrimsson, J. A., Hanley, D. F., and Rosenblum, M. (2017). Improving precision by adjusting for prognostic baseline variables in randomized trials with binary outcomes, without regression model assumptions. *Contemporary Clinical Trials*, 54:18–24.

Ta, T., Shao, J., Li, Q., and Wang, L. (2020). Generalized regression estimators with high-dimensional covariates. *Statistica Sinica*, 30(3):1135–1154.

Taves, D. R. (1974). Minimization: A new method of assigning patients to treatment and control groups. *Clinical Pharmacology and Therapeutics*, 15(5):443–453.

Tsiatis, A. A., Davidian, M., Zhang, M., and Lu, X. (2008). Covariate adjustment for two-sample treatment comparisons in randomized clinical trials: A principled yet flexible approach. *Statistics in Medicine*, 27(23):4658–4677.

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

Wang, B., Susukida, R., Mojtabai, 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*.

White, H. (1980). A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity. *Econometrica*, 48(4):817–838.

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

Yates, M., Mootoo, A., Adas, M., Bechman, K., Rampes, S., Patel, V., Qureshi, S., Cope, A. P., Norton, S., and Galloway, J. B. (2021). Venous thromboembolism risk with jak inhibitors: A meta-analysis. *Arthritis & Rheumatology*, 73(5):779–788.

Ye, T. (2018). Testing hypotheses under covariate-adaptive randomisation and additive models. *Statistical Theory and Related Fields*, 2(1):96–101.

Ye, T., Yi, Y., and Shao, J. (2020). Inference on average treatment effect under minimization and other covariate-adaptive randomization methods. *arXiv preprint arXiv:2007.09576*.

Zelen, M. (1974). The randomization and stratification of patients to clinical trials. *Journal of Chronic Diseases*, 27(7):365–375.

Table 1: Bias, standard deviation (SD), average standard error (SE), and coverage probability (CP) of 95% asymptotic confidence interval for linear contrasts under simple randomization (SR), stratified permuted block randomization (PB), and Pocock-Simon's minimization based on 10,000 simulations and setup in (21) with $n = 400$

				$\theta_2 - \theta_1$				$\theta_3 - \theta_1$			
Allocation	Randomization	Method	X	Bias	SD	SE	CP	Bias	SD	SE	CP
1:1:1	SR	ANOVA		0.001	0.327	0.327	0.946	-0.001	0.201	0.201	0.949
		ANCOVA	Z	0.001	0.310	0.306	0.945	0.000	0.229	0.227	0.949
		ANCOVA	Z, U, W	-0.011	0.295	0.289	0.944	0.003	0.246	0.240	0.945
		ANHECOVA	Z	-0.001	0.271	0.265	0.944	-0.002	0.169	0.167	0.949
		ANHECOVA	Z, U, W	0.000	0.226	0.224	0.947	-0.002	0.155	0.152	0.948
	PB	ANOVA		0.000	0.267	0.327	0.982	-0.001	0.166	0.201	0.982
		ANCOVA	Z	0.001	0.267	0.306	0.975	-0.001	0.166	0.227	0.992
		ANCOVA	Z, U, W	-0.012	0.250	0.289	0.974	0.003	0.186	0.240	0.987
		ANHECOVA	Z	0.001	0.266	0.265	0.947	-0.001	0.165	0.167	0.950
		ANHECOVA	Z, U, W	0.002	0.224	0.224	0.950	-0.001	0.152	0.152	0.948
	Minimization	ANOVA		-0.002	0.268	0.327	0.982	0.000	0.168	0.201	0.980
		ANCOVA	Z	-0.002	0.267	0.306	0.974	0.000	0.168	0.227	0.992
		ANCOVA	Z, U, W	-0.013	0.250	0.289	0.974	0.005	0.189	0.240	0.987
		ANHECOVA	Z	-0.002	0.267	0.265	0.948	0.000	0.167	0.167	0.949
		ANHECOVA	Z, U, W	-0.001	0.225	0.223	0.946	0.000	0.154	0.152	0.944
1:2:2	SR	ANOVA		0.001	0.311	0.312	0.951	0.000	0.205	0.204	0.951
		ANCOVA	Z	0.001	0.298	0.294	0.947	0.001	0.241	0.239	0.950
		ANCOVA	Z, U, W	-0.013	0.286	0.281	0.945	0.003	0.262	0.257	0.947
		ANHECOVA	Z	0.000	0.268	0.268	0.949	0.001	0.182	0.180	0.948
		ANHECOVA	Z, U, W	0.001	0.237	0.238	0.949	0.002	0.171	0.169	0.944
	PB	ANOVA		0.002	0.265	0.312	0.977	0.001	0.180	0.204	0.975
		ANCOVA	Z	0.002	0.265	0.293	0.969	0.001	0.181	0.238	0.991
		ANCOVA	Z, U, W	-0.012	0.253	0.280	0.967	0.005	0.207	0.256	0.986
		ANHECOVA	Z	0.002	0.264	0.267	0.950	0.001	0.179	0.179	0.949
		ANHECOVA	Z, U, W	0.001	0.234	0.238	0.951	0.002	0.170	0.168	0.946
	Minimization	ANOVA		0.000	0.265	0.311	0.979	0.002	0.181	0.204	0.971
		ANCOVA	Z	-0.001	0.265	0.293	0.969	0.001	0.182	0.238	0.991
		ANCOVA	Z, U, W	-0.013	0.253	0.280	0.970	0.006	0.208	0.256	0.985
		ANHECOVA	Z	-0.001	0.265	0.267	0.951	0.001	0.181	0.179	0.947
		ANHECOVA	Z, U, W	0.001	0.233	0.238	0.956	0.002	0.172	0.168	0.945

Table 2: Bias, standard deviation (SD), average standard error (SE), and coverage probability (CP) of 95% asymptotic confidence interval for ratios under simple randomization (SR), stratified permuted block randomization (PB), and Pocock-Simon's minimization based on 10,000 simulations and setup in (21) with $n = 400$

				θ_2/θ_1				θ_3/θ_1			
Allocation	Randomization	Method	X	Bias	SD	SE	CP	Bias	SD	SE	CP
1:1:1	SR	ANOVA		0.022	0.379	0.378	0.955	0.020	0.259	0.258	0.952
		ANCOVA	Z	0.022	0.368	0.362	0.950	0.023	0.296	0.292	0.949
		ANCOVA	Z, U, W	0.039	0.360	0.352	0.950	0.027	0.321	0.313	0.946
		ANHECOVA	Z	0.024	0.337	0.329	0.951	0.020	0.232	0.228	0.948
		ANHECOVA	Z, U, W	0.022	0.304	0.298	0.948	0.019	0.221	0.215	0.947
	PB	ANOVA		0.021	0.331	0.377	0.978	0.018	0.228	0.257	0.973
		ANCOVA	Z	0.021	0.331	0.359	0.972	0.018	0.228	0.289	0.987
		ANCOVA	Z, U, W	0.038	0.324	0.349	0.970	0.020	0.256	0.310	0.979
		ANHECOVA	Z	0.020	0.331	0.328	0.954	0.018	0.227	0.227	0.952
		ANHECOVA	Z, U, W	0.020	0.301	0.297	0.951	0.017	0.217	0.214	0.947
	Minimization	ANOVA		0.026	0.327	0.378	0.981	0.021	0.232	0.258	0.972
		ANCOVA	Z	0.026	0.327	0.360	0.974	0.021	0.233	0.290	0.985
		ANCOVA	Z, U, W	0.041	0.321	0.350	0.972	0.022	0.262	0.310	0.980
		ANHECOVA	Z	0.026	0.326	0.329	0.958	0.021	0.231	0.228	0.951
		ANHECOVA	Z, U, W	0.026	0.299	0.297	0.953	0.021	0.220	0.215	0.948
1:2:2	SR	ANOVA		0.035	0.410	0.405	0.955	0.030	0.303	0.298	0.947
		ANCOVA	Z	0.042	0.422	0.409	0.949	0.042	0.366	0.353	0.947
		ANCOVA	Z, U, W	0.070	0.435	0.419	0.948	0.054	0.406	0.389	0.945
		ANHECOVA	Z	0.036	0.383	0.373	0.950	0.029	0.285	0.277	0.942
		ANHECOVA	Z, U, W	0.034	0.362	0.352	0.948	0.027	0.278	0.266	0.940
	PB	ANOVA		0.033	0.373	0.403	0.969	0.028	0.279	0.296	0.962
		ANCOVA	Z	0.033	0.374	0.401	0.967	0.028	0.280	0.344	0.981
		ANCOVA	Z, U, W	0.059	0.386	0.410	0.965	0.038	0.326	0.379	0.974
		ANHECOVA	Z	0.034	0.373	0.372	0.953	0.028	0.279	0.275	0.948
		ANHECOVA	Z, U, W	0.034	0.354	0.350	0.949	0.027	0.272	0.265	0.945
	Minimization	ANOVA		0.030	0.372	0.401	0.967	0.025	0.276	0.293	0.961
		ANCOVA	Z	0.039	0.374	0.401	0.966	0.033	0.281	0.344	0.982
		ANCOVA	Z, U, W	0.062	0.390	0.408	0.963	0.040	0.330	0.377	0.973
		ANHECOVA	Z	0.035	0.372	0.371	0.952	0.029	0.276	0.274	0.950
		ANHECOVA	Z, U, W	0.036	0.352	0.349	0.950	0.028	0.269	0.264	0.947

Table 3: Estimate, standard error (SE), and p-value in the real data example analysis

		Physician versus placebo			Soccer star versus placebo		
Method	X	Difference	SE	p-value	Difference	SE	p-value
ANOVA		0.386	0.211	0.067	-0.068	0.205	0.739
ANCOVA	Grade	0.403	0.203	0.046	-0.052	0.203	0.799
	Grade, Anemia status	0.437	0.199	0.028	-0.085	0.201	0.672
ANHECOVA	Grade	0.409	0.200	0.041	-0.051	0.201	0.800
	Grade, Anemia status	0.481	0.193	0.013	-0.046	0.195	0.815
		Ratio	SE	p-value	Ratio	SE	p-value
ANOVA		1.034	0.018	0.062	0.994	0.019	0.752
ANCOVA	Grade	1.035	0.018	0.051	0.996	0.018	0.800
	Grade, Anemia status	1.038	0.018	0.033	0.993	0.017	0.670
ANHECOVA	Grade	1.036	0.018	0.045	0.996	0.018	0.800
	Grade, Anemia status	1.042	0.017	0.016	0.996	0.017	0.803

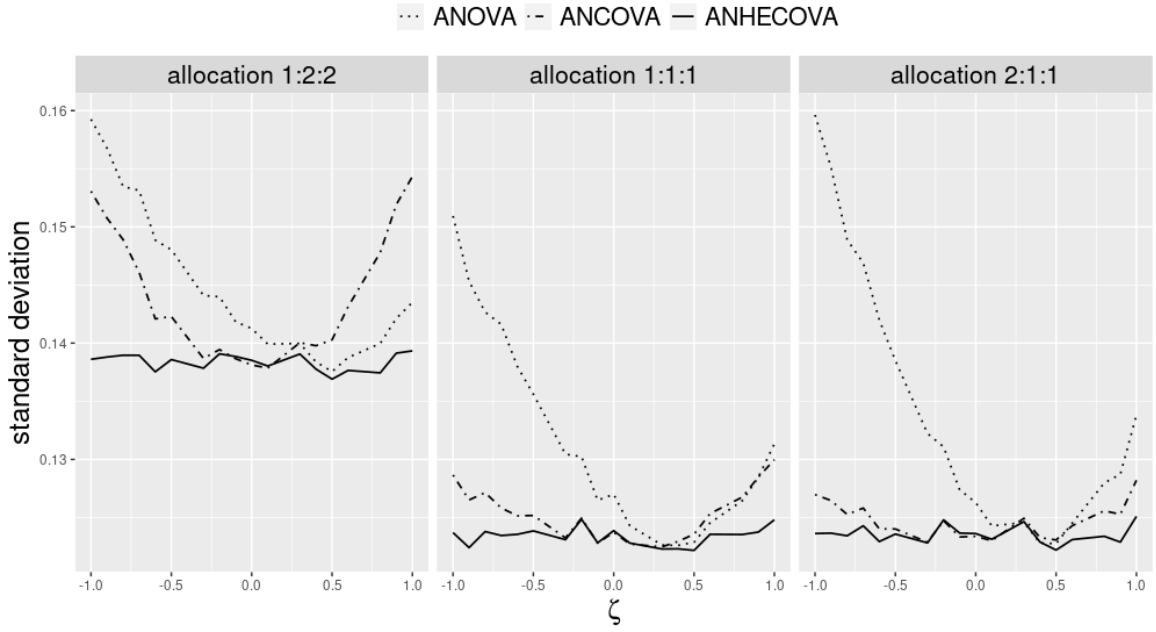

Figure 1: Standard deviations of ANOVA, ANCOVA, and ANHECOVA estimators of $\theta_2 - \theta_1$ based on 10,000 simulations and setup in (20)