Regression analysis for covariate-adaptive randomization: A robust and efficient inference perspective

Wei Ma\(^1\) | Fuyi Tu\(^1\) | Hanzhong Liu\(^2\)

\(^1\)Institute of Statistics and Big Data, Renmin University of China, Beijing, China
\(^2\)Center for Statistical Science, Department of Industrial Engineering, Tsinghua University, Beijing, China

Correspondence
Hanzhong Liu, Center for Statistical Science, Department of Industrial Engineering, Tsinghua University, Beijing, China.
Email: lhz2016@tsinghua.edu.cn

Funding information
National Key R&D Program of China, Grant/Award Number: 2018YFC2000302; National Natural Science Foundation of China, Grant/Award Numbers: 12071242, 12171476; Guo Qiang Institute of Tsinghua University, Grant/Award Number: 2020GQG1011

Linear regression is arguably the most fundamental statistical model; however, the validity of its use in randomized clinical trials, despite being common practice, has never been crystal clear, particularly when stratified or covariate-adaptive randomization is used. In this article, we investigate several of the most intuitive and commonly used regression models for estimating and inferring the treatment effect in randomized clinical trials. By allowing the regression model to be arbitrarily misspecified, we demonstrate that all these regression-based estimators robustly estimate the treatment effect, albeit with possibly different efficiency. We also propose consistent non-parametric variance estimators and compare their performances to those of the model-based variance estimators that are readily available in standard statistical software. Based on the results and taking into account both theoretical efficiency and practical feasibility, we make recommendations for the effective use of regression under various scenarios. For equal allocation, it suffices to use the regression adjustment for the stratum covariates and additional baseline covariates, if available, with the usual ordinary-least-squares variance estimator. For unequal allocation, regression with treatment-by-covariate interactions should be used, together with our proposed variance estimators. These recommendations apply to simple and stratified randomization, and minimization, among others. We hope this work helps to clarify and promote the usage of regression in randomized clinical trials.

KEYWORDS
ANCOVA, covariate-adaptive randomization, minimization, regression adjustment, stratified randomization

1 | INTRODUCTION

The use of linear regression models is standard practice for estimating the treatment effect in randomized clinical trials, regardless of whether simple randomization, a covariate-adaptive allocation, or other approaches are used. However, concerns have been raised regarding the validity of the resulting inferences, because the “usual” assumptions for linear regression, such as linearity, normality, and homoskedasticity, may not be fulfilled, and certainly cannot be verified during the development of a statistical analysis plan prior to initiation of a trial. Consequently, the use of regression in randomized clinical trials must be justified by demonstrating that valid inference can be obtained even when the regression model is
arbitrarily misspecified. Although this is not the case in general (e.g., in observational studies), regression does exhibit the desired robustness thanks to randomization.

The robustness of regression is evident under simple randomization. The seminal work of Yang and Tsiatis examined three commonly used regression models, namely, simple regression, which is equivalent to the difference-in-means estimator, and two regression models that adjust for covariates with and without covariate-by-treatment interactions. The authors proved that all three ordinary-least-squares (OLS) estimators are consistent and they argued that the regression with covariate-by-treatment interactions produces the most efficient estimator. From a practical standpoint, Wang et al. pointed out that the OLS variance estimator is consistent under equal allocation for regression without interactions, which supports the use of standard statistical software (e.g., lm in R or proc reg in SAS) for constructing confidence intervals and tests.

Similar results have also been obtained under finite population framework. Freedman, being apparently unaware of the work by Yang and Tsiatis, criticized regression because the adjustment (without interactions) can actually reduce the asymptotic precision compared with that obtained using the difference-in-means estimator. The issue was later addressed in the influential paper by Lin: the regression with interactions cannot hurt, and often improves, the asymptotic precision. Lin also showed that the Huber-White variance estimator is either consistent or asymptotically conservative.

Compared with simple randomization, the robustness of regression in stratified randomization, or more generally, covariate-adaptive randomization, appears to be more elusive, despite its extensive use in clinical trials. Covariate-adaptive randomization aims to balance treatment allocation among covariates. For example, stratified randomization defines a set of strata based on one or more covariates and performs a separate randomization, most commonly block randomization, within each stratum. Pocock and Simon proposed a minimization procedure, which we refer to simply as minimization in this article, to achieve balance over covariates’ margins. This approach has been generalized for use in simultaneously controlling various types of imbalances (within-stratum, within-covariate-margin, and overall). For a more comprehensive review, see Rosenberger. The use of covariate-adaptive randomization predominates in clinical trials. According to a recent survey of 224 randomized trials published in leading medical journals in 2014, over 80% of them had used covariate-adaptive randomization.

The past decade has witnessed significant advances in inference under covariate-adaptive randomization. Various valid tests have been proposed using different randomization methods and model assumptions. However, one drawback of most of these studies is that the data generation model must be correctly specified. More recently, the topic of robust inference under covariate-adaptive randomization has drawn much attention and has been tackled from different perspectives such as regression adjustment, M-estimation, and survival analysis. Most notably, Bugni et al. studied the two-sample t test and the regression that adjusts for stratification indicators under stratified randomization, without imposing specific functional form requirements on the data generation model. Under finite population framework, Liu and Yang also proposed a regression-adjusted treatment effect estimator under stratified block randomization.

In addition to the stratum covariates, additional baseline covariates may be available. Generally, the appropriate adjustment of baseline covariates can improve statistical efficiency. However, although the stratification indicators are routinely adjusted to analyze clinical trial data as required by regulatory guidelines, the inferential properties remain largely unknown when additional covariates are adjusted in the regression analysis. As pointed out in Wang et al., “It is an open question, to the best of our knowledge, as to what happens when more variables than the stratification indicators are included in the ANCOVA model under such randomization schemes (covariate-adaptive randomization), in terms of consistency of the ANCOVA estimator and how to compute its asymptotic variance under arbitrary model misspecification.”

Because covariate-adaptive randomization is so common in clinical trials, it is crucial to clarify and provide guidance on the use of regression models for covariate-adaptive randomized trials, with the acknowledgment that the models may be misspecified and that additional baseline covariates may be available. In particular, the following two questions demand definite answers: (i) Which linear regression model can be used to ensure robust and efficient estimation of the treatment effect? (ii) Is a consistent variance estimator available and easily accessible to facilitate valid inference?

This article seeks to address these questions. We evaluate six commonly used regression models for estimating and inferring treatment effects under a broad range of covariate-adaptive randomizations. These models cover the scenarios in which only stratum covariates are used for analysis as well as when additional baseline covariates are also adjusted. Based on the asymptotic results, the most efficient estimator is identified. Moreover, we propose non-parametric consistent variance estimators for all of the regression estimators under consideration. We also examine the OLS variance estimator, which is commonly used in practice, and indicate when it can be used for valid inference. Finally, we make
practical recommendations for the use of regression under covariate-adaptive randomization. All of the regression methods considered in this article are implemented in the R package caratREG, available at GitHub (https://github.com/FuyiTu/caratREG).

2 | FRAMEWORK AND NOTATION

Consider a covariate-adaptive randomized experiment with $n$ units. For each unit $i = 1, \ldots, n$, let $A_i$ denote the treatment assignment with $A_i = 1$ for the treatment and $A_i = 0$ for the control, and $X_i$ denote a $p$-dimensional column vector of baseline covariates. Let $n_1 = \sum_{i=1}^{n} A_i$ and $n_0 = \sum_{i=1}^{n} (1 - A_i)$ denote the number of treated and control units, respectively. The experimental units are stratified into $K$ strata based on baseline covariates $X_i$ using a function $B$: $\text{supp}(X_i) \rightarrow \{1, \ldots, K\}$, where $\text{supp}(X_i)$ is the support of $X_i$. For example, the units are stratified according to gender, grade or location. For simplicity, we assume that units are assigned to each stratum with positive probability, that is, $p_i[k] = P(B_i = k) > 0$, $i = 1, \ldots, n$, $k = 1, \ldots, K$, where $B_i = B(X_i)$ is the stratum label. In stratum $k$, let $[k]$ be the index set of units, let $n_{[k]} = \sum_{i \in [k]} 1$, $n_{[k]} = \sum_{i \in [k]} A_i$, $n_{[k]} = \sum_{i \in [k]} (1 - A_i)$ denote the number of units, the number of treated units, and the number of control units, respectively, and let $p_{[k]} = n_{[k]}/n$ and $\pi_{[k]} = n_{[k]}/n_{[k]}$ denote the proportions of stratum sizes and the proportions of treated units, respectively. The theoretical property of covariate-adaptive randomization critically depends on the difference between the actual treatment assignment and the target treatment proportion $\pi \in (0, 1)$:

$$D_{[k]} = \sum_{i=1}^{n} (A_i - \pi) I_{i \in [k]}, \quad k = 1, \ldots, K,$$

where $I_{i \in [k]}$ is an indicator function which equals one if $i \in [k]$ and zero otherwise.

We use the Neyman-Rubin potential outcomes model\cite{22,23} to define the treatment effect. For unit $i$, this model assumes that there are two potential outcomes, $Y_i(1)$ and $Y_i(0)$, when exposed to the treatment and to the control, respectively. Under the stable unit treatment value assumption,\cite{24} the observed outcome $Y_i$ is a function of the treatment assignment and the potential outcomes:

$$Y_i = A_i Y_i(1) + (1 - A_i) Y_i(0).$$

Our goal is to estimate and infer the treatment effect $\tau = E\{Y_i(1) - Y_i(0)\}$ based on the observed data $\{Y_i, A_i, B_i, X_i, i = 1, \ldots, n\}$.

We make the following assumptions regarding the data generation process and covariate-adaptive randomization. We denote the set of univariate random variables with at least one positive stratum-specific variance as $R_2 = \{V : \max_{k=1,\ldots,K} \text{Var}(V|B_i = k) > 0\}$. Let $A^{(n)} = \{A_1, \ldots, A_n\}$, $B^{(n)} = \{B_1, \ldots, B_n\}$, and $W^{(n)} = \{W_1, \ldots, W_n\}$, where $W_i = (Y_i(1), Y_i(0), X^T_i)$, $i = 1, \ldots, n$, are independent and identically distributed (i.i.d.) samples from the population distribution of $W = (Y(1), Y(0), X^T)^T$.

**Assumption 1.** $E\{|Y_i(a)|^2\} < \infty$ and $Y_i(a) \in R_2$ for $a = 0, 1$.

**Assumption 2.** $W^{(n)} \perp A^{(n)} | B^{(n)}$.

**Assumption 3.** $\{n^{-1/2} D_{[k]} k = 1, \ldots, K \} \xrightarrow{d} \mathcal{N}(0, \Sigma_D)$ a.s., where $\Sigma_D = \text{diag}\{p_{[k]} q_{[k]} : k = 1, \ldots, K\}$ with $0 \leq q_{[k]} \leq \pi(1 - \pi)$ for $k = 1, \ldots, K$.

**Assumption 4.** $D_{[k]} = O_p(\sqrt{n})$ for $k = 1, \ldots, K$.

Assumption 2 requires that given the stratum covariates, the treatment assignments are conditionally independent of the potential outcomes and additional covariates not used in the randomization. This assumption is satisfied for covariate-adaptive randomization for which the treatment assignments depend only on the vector of stratum covariates. Assumption 3 was proposed by Bugni et al\cite{15} to study statistical inference under covariate-adaptive randomization with the same target treatment proportion $\pi$ across all strata. This assumption characterizes the asymptotic behavior of jointly independent imbalances within strata and thus is particularly relevant to stratified randomization. As simple randomization leads to $q_{[k]} = \pi(1 - \pi)$, it is reasonable to expect that a smaller value of $q_{[k]}$ can be achieved by stratified randomization. For example, when used within each stratum, Wei’s urn design\cite{25} leads to a $q_{[k]}$ value between zero
and \( \pi(1 - \pi) \), whereas block randomization\(^{26} \) and Efron’s biased coin design and its generalizations\(^{27,28} \) can reduce \( q_{[k]} \) to zero. In the latter case of \( q_{[k]} = 0 \), we say that the randomization achieves strong balance.\(^{15} \) Moreover, Assumption 4 is weaker than Assumption 3 as no (asymptotic) independence is required between different strata, and it allows us to consider covariate-adaptive randomization that pursues balance within covariate margins and thus has a complicated dependence structure across strata, such as minimization\(^{5,7} \) and the class of designs proposed by Hu and Hu.\(^6 \) We emphasize that while Assumption 3 applies primarily to stratified randomization, Assumption 4 is quite general and is satisfied by most covariate-adaptive randomization methods encountered in practice.

### 3 | ADJUSTMENT FOR STRATUM COVARIATES

#### 3.1 | Difference-in-means

Using the difference in the sample means is perhaps the most straightforward and intuitive approach for estimating the treatment effect. Unadjusted methods such as this, which also include the two-sample \( t \) test, are commonly used for analyzing data from randomized clinical trials due to their simplicity, transparency, and robustness to model-misspecification, among other practical reasons.\(^{4,10} \) Although failure to adjust baseline covariates may reduce the efficiency and is contrary to regulatory guidelines for analyzing clinical trial data with covariate-adaptive randomization,\(^{20,21} \) we start with the difference-in-means estimator because of its popularity in practice and its theoretical importance in its own right. It also serves as a benchmark for evaluating subsequent estimators.

Denote the sample means of the observed outcomes for the treatment and control groups as \( \bar{Y}_1 = (1/n_1) \sum_{i=1}^{n_1} A_i Y_i \) and \( \bar{Y}_0 = (1/n_0) \sum_{i=1}^{n_0} (1 - A_i) Y_i \), respectively. The difference-in-means estimator \( \hat{\tau} = \bar{Y}_1 - \bar{Y}_0 \) is equal to the OLS estimator of the coefficient of \( A_i \) in the following regression:

\[
Y_i \sim \alpha + A_i \tau.
\]

In this article, we use the notation “\( \sim \)” to indicate that the regression coefficients are obtained by regressing \( Y_i \) on the quantities on the right hand side (on the sample level). For example, the above expression means regressing \( Y_i \) on \( A_i \) with intercept.

To state the asymptotic results of \( \hat{\tau} \) (and two alternative estimators introduced shortly in this section), we use the following notations due to Bugnii et al.\(^{15} \)

\[
\begin{align*}
\varsigma_Y^2(\pi) &= \frac{1}{\pi} \text{Var}\{\bar{Y}_1(1)\} + \frac{1}{1 - \pi} \text{Var}\{\bar{Y}_0(0)\}, \\
\varsigma_H^2 &= E\{\bar{Y}_1(1) - \bar{Y}_0(0)\}^2, \\
\varsigma_A^2(\pi) &= E[q_{[B]} \left\{ \frac{1}{\pi} \bar{Y}_1(1) + \frac{1}{1 - \pi} \bar{Y}_0(0) \right\}]^2, \\
\varsigma_\pi^2 &= \frac{(1 - 2\pi)^2}{\pi^2(1 - \pi)^2} E[q_{[B]} \{\bar{Y}_1(1) - \bar{Y}_0(0)\}]^2,
\end{align*}
\]

where \( \bar{Y}_1(a) = Y_1(a) - E[Y_1(a)|B_1] \) and \( \bar{Y}_0(a) = E[Y_1(a)|B_1] - E[Y_1(a)], \) \( a = 0, 1. \) We denote their sample analogs as \( \hat{\varsigma}^2_Y(\pi), \hat{\varsigma}^2_H, \hat{\varsigma}^2_A(\pi), \) and \( \hat{\varsigma}^2_\pi, \) respectively, and the detailed expressions are given in the Supplementary Material. Subsequently, we have the following proposition.\(^{15} \)

**Proposition 1.** Under Assumptions 1–3,

\[
\sqrt{n}(\hat{\tau} - \tau) \xrightarrow{d} \mathcal{N}(0, \varsigma_Y^2(\pi) + \varsigma_H^2 + \varsigma_A^2(\pi)),
\]

and the asymptotic variance can be consistently estimated by \( \hat{\varsigma}_Y^2(\pi) + \hat{\varsigma}_H^2 + \hat{\varsigma}_A^2(\pi). \)

#### 3.2 | Regression adjustment without interaction

It is generally recognized that adjusting for baseline covariates can help to remedy the imperfect balance with respect to the covariates in a completely randomized experiment and thus improve the precision in estimating the treatment effect.\(^{29,30} \) As noted in the previous section, regulatory guidelines also recommend that the covariates used in the
randomization be adjusted in the subsequent analysis. The most basic way to do so is to include the stratification indicators in the linear regression model, that is,

\[ Y_i \sim \alpha + A_i \tau + \sum_{k=1}^{K-1} a_k I_{i \in [k]}, \]  

(1)

Let \( \hat{\tau}_{\text{adj}} \) and \( \hat{\sigma}_2 \) be the OLS point and variance estimators of \( \tau \). Denote the stratum-specific treatment effect in stratum \( k \) \( (k = 1, \ldots, K) \) as \( \tau_{[k]} = E[Y_i | (1 - Y_i)B_i = 1] \) and its difference-in-means estimator as \( \hat{\tau}_{[k]} = \bar{Y}_{[k1]} - \bar{Y}_{[k0]} \), where \( \bar{Y}_{[k]} = (1/n_{[k]}) \sum_{i \in [k]} Y_i \) and \( \bar{Y}_{[k0]} = (1/n_{[k0]}) \sum_{i \in [k]} (1 - A_i) Y_i \) are the stratum-specific sample means for the treatment and control groups, respectively. Another equivalent and useful formula of \( \hat{\tau}_{\text{adj}} \) given by Bugnietal \(^{15}\) is

\[ \hat{\tau}_{\text{adj}} = \sum_{k=1}^{K} \omega_{[k]} \hat{\tau}_{[k]}, \quad \omega_{[k]} = \frac{\pi n_{[k]}(1 - \pi n_{[k]}) \rho_{[k]} (\pi n_{[k]} - \pi n_{[k']}) \rho_{[k']}}{\sum_{k'=1}^{K} \pi n_{[k']} (1 - \pi n_{[k']} \rho_{[k']} \rho_{[k']}).} \]

**Proposition 2.** Under Assumptions 1–3,

\[ \sqrt{n} (\hat{\tau}_{\text{adj}} - \tau) \xrightarrow{d} \mathcal{N}(0, \psi^2(\pi) + \psi^2_H + \psi^2_\pi), \quad n\hat{\sigma}^2_{\text{adj}} \xrightarrow{p} \psi^2(1 - \pi) + \psi^2_H. \]

and the asymptotic variance can be consistently estimated by \( \psi^2(\pi) + \psi^2_H + \psi^2_\pi \). Furthermore, when \( \pi = 1/2 \), the conclusions hold if Assumption 3 is replaced by Assumption 4.

Bugniet.al \(^{15}\) obtained the asymptotic normality of \( \hat{\tau}_{\text{adj}} \), the consistent variance estimator, and the limit of the OLS variance estimator under Assumptions 1–3. The important observation that Assumption 3 can be weakened to Assumption 4 when \( \pi = 1/2 \) is a new result that has not appeared in the literature to the best of our knowledge.

The performances of \( \hat{\tau}_{\text{adj}} \) and \( \hat{\sigma}^2 \) depend on whether the target treatment allocation is balanced. For equal allocation \( (\pi = 1/2) \), because \( \psi^2_\pi = 0 \) and \( \psi^2_H(\pi) \geq 0 \), the asymptotic variance of \( \hat{\tau}_{\text{adj}} \) is less than or equal to that of \( \hat{\tau} \). That is, the regression improves, or at least does not hurt, the precision of estimating and inferring the treatment effect \( \tau \). Moreover, \( \psi^2_\pi = 0 \) and \( \psi^2_H(\pi) = \psi^2_H(1 - \pi) \) imply that \( \hat{\sigma}^2 \) is a consistent estimator for the asymptotic variance of \( \hat{\tau}_{\text{adj}} \). More importantly, Proposition 2 holds under Assumption 4. Thus, for most covariate-adaptive randomized trials, it is valid to use the usual OLS point and variance estimators obtained using regression without interactions. The validity of the usual OLS point and variance estimators was confirmed by Yang and Tsai\(^{1}\) and Wang et al.\(^{2}\) but under simple randomization.

For unequal allocation \( (\pi \neq 1/2) \), \( \hat{\sigma}^2 \) can be anti-conservative even for a strong balance treatment assignment, which results in an inflated type I error rate. Even if we use the consistent variance estimator \( \psi^2(\pi) + \psi^2_H + \psi^2_\pi \) to construct a valid confidence interval or test, we should keep in mind that the performance of \( \hat{\tau}_{\text{adj}} \) can be worse than that of the difference-in-means estimator.

### 3.3 Regression adjustment with interaction

In this section, we study the regression model with stratification-by-treatment interactions under covariate-adaptive randomization. This is motivated in part by results reported in the literature, which state that adding the interaction terms can further improve precision when estimating the treatment effect under various randomization methods.\(^{1,4,16,19}\)

The third regression under consideration is

\[ Y_i \sim \alpha + A_i \tau + \sum_{k=1}^{K-1} a_k I_{i \in [k]} + \sum_{k=1}^{K-1} v_k A_i (I_{i \in [k]} - p_{n_{[k]} \pi_{[k]}}). \]

(2)

The OLS estimator of \( \tau \), the coefficient of \( A_i \), is denoted as \( \hat{\tau}_{\text{interact}} \). In this case, we center the stratum indicator \( I_{i \in [k]} \) at its sample mean \( p_{n_{[k]} \pi_{[k]}} \). In the interactions to ensure that \( \hat{\tau}_{\text{interact}} \) alone can be used to estimate the treatment effect \( \tau \) without referring to a linear combination of multiple estimated regression coefficients. This approach is equivalent to but more convenient than the fully saturated regression estimator obtained by a two-step procedure in Bugniet al.\(^{16}\) In fact, \( \hat{\tau}_{\text{interact}} \) has a more intuitive expression that can be considered as a stratified difference-in-means estimator.\(^{19}\)
\[ \hat{\tau}_{\text{interact}} = \sum_{k=1}^{K} p_{n[k]}(\bar{Y}_{[k]} - \bar{Y}_{[k']}) = \sum_{k=1}^{K} p_{n[k]} \hat{\tau}_{[k]}. \]

As \( \tau = \sum_{k=1}^{K} p_{[k]} \hat{\tau}_{[k]} \), \( \hat{\tau}_{\text{interact}} \) is a natural plug-in estimator. Let \( \hat{\sigma}^2_{\text{interact}} \) be the OLS variance estimator of \( \hat{\tau}_{\text{interact}} \).

**Proposition 3.** Under Assumptions 1, 2, and 4,

\[ \frac{\sqrt{n}(\hat{\tau}_{\text{interact}} - \tau)}{\hat{\sigma}_{\text{interact}}} \xrightarrow{d} N(0, \varphi^2_\tau + \varphi^2_H), \quad \frac{n\hat{\sigma}^2_{\text{interact}}}{\varphi^2_\tau} \xrightarrow{p} \varphi^2_\tau(1 - \pi). \]

and the asymptotic variance can be consistently estimated by \( \varphi^2_\tau(\pi) + \varphi^2_H \).

**Remark 1.** The results in the proposition were obtained by Bugni et al.\(^{16}\) under a weaker condition allowing the target treatment proportion to differ across strata. Although this setting is uncommon in clinical trials, it could be useful in other fields. We refer interested readers to Bugni et al.\(^{16}\) for more details. See also Remark 4.

Based on the asymptotic normality and using the consistent variance estimator, we can construct a valid confidence interval or test for \( \tau \). Moreover, Proposition 3 implies that the OLS variance estimator \( \hat{\sigma}^2_{\text{interact}} \) in regression (2) can be anti-conservative even when \( \tau = 1/2 \), so we should not use it. As in many clinical trials, \( \tau = 1/2 \), we discuss this special case in more details.

**Corollary 1.** When \( \tau = 1/2 \), under Assumptions 1, 2 and 4, \( \hat{\tau}_{\text{interact}} \) and \( \hat{\tau}_{\text{adj}} \) are asymptotically equivalent, with both asymptotic variances being smaller than or equal to that of \( \hat{\tau} \). Furthermore, both \( n\hat{\sigma}^2_{\text{adj}} \) and \( \varphi^2_\tau(\pi) + \varphi^2_H \) are consistent variance estimators.

Corollary 1 implies that for almost all covariate-adaptive randomization with equal allocation, it suffices to use regression without interactions, that is, regression (1), to estimate and infer \( \tau \) when only stratum covariates are adjusted in the analysis.

### 3.4 Summary and recommendation

The asymptotic variances and variance estimators of \( \hat{\tau} \), \( \hat{\tau}_{\text{adj}} \), and \( \hat{\tau}_{\text{interact}} \) are summarized in Table 1. As \( \varphi^2_\tau(\pi) \) and \( \varphi^2_\pi \) are non-decreasing functions of \( q[\pi] \), \( \hat{\tau} \) and \( \hat{\tau}_{\text{adj}} \) under covariate-adaptive randomization are always no worse than those obtained under simple randomization. Thus, covariate-adaptive randomization can improve efficiency compared to simple randomization, even when adjusting for stratification indicators in the analysis. Generally, the asymptotic variances of \( \hat{\tau} \) and \( \hat{\tau}_{\text{adj}} \) are not ordered unambiguously when \( \pi \neq 1/2 \). Moreover, as \( \varphi^2_\tau \geq 0 \) and \( \varphi^2_\pi(\pi) \geq 0 \), the stratified difference-in-means estimator \( \hat{\tau}_{\text{interact}} \) has the smallest asymptotic variance. When \( \pi = 1/2 \), it holds that \( \varphi^2_\tau = 0 \), thus \( \hat{\tau}_{\text{adj}} \) and \( \hat{\tau}_{\text{interact}} \) are asymptotically equivalent, both being no worse than \( \hat{\tau} \). Furthermore, when the treatment assignment achieves strong balance (\( q[\pi] = 0 \), \( k = 1, \ldots, K \)), \( \varphi^2_\tau(\pi) = \varphi^2_\pi = 0 \), which implies that \( \hat{\tau} \), \( \hat{\tau}_{\text{adj}} \), and \( \hat{\tau}_{\text{interact}} \) are asymptotically equivalent. Taking into account both simplicity and efficiency, our recommendations are summarized in Table 2. Note that, the asymptotic variances of \( \hat{\tau}_{\text{interact}} \) and, when \( \pi = 1/2 \), \( \hat{\tau}_{\text{adj}} \) do not depend on a particular covariate-adaptive randomization procedure as long as the randomization satisfies Assumption 4. Thus, the recommendations apply for most commonly used covariate-adaptive randomization methods, including stratified randomization and minimization.

### 4 Adjustment for Additional Baseline Covariates

Apart from the stratification indicators, the covariates \( X_i \) may contain additional information with which to better estimate the treatment effect. The additional covariates may include not only those excluded from the design stage but also the continuous covariates that have been discretized for use in covariate-adaptive randomization. Regression adjustment is a common strategy for adjusting for additional covariates to improve the precision.\(^{19,31}\) In the following, we discuss three widely used regression methods that adjust for additional covariates, which correspond to the three regressions discussed in Section 3 with the addition of \( X_i \). Note that if a discrete covariate has already been adjusted in the regression as a stratification indicator, it is not, per se, an additional covariate. In general, we should remove the covariates that can be linearly represented by the stratification indicators from \( X_i \). For simplicity, however, we continue to use the same notation.
TABLE 1 Summary of asymptotic results for various regression-based treatment effect estimators

| Regression adjustment | Target allocation | Estimator | Asymptotic variance | Proposed consistent variance estimator | Is OLS variance estimator valid? |
|-----------------------|------------------|-----------|---------------------|----------------------------------------|-------------------------------|
| Stratification only   | 1/2              | \( \hat{r} \) | \( \sigma_r^2(\pi) + \sigma_{Hr}^2(\pi) \) | \( \sigma_r^2(\pi) + \sigma_{Hr}^2(\pi) \) | No |
| #1/2                  |                  | \( \hat{r}_{\text{adj}} \) | \( \sigma_r^2(\pi) + \sigma_{Hr}^2(\pi) \) | \( \sigma_r^2(\pi) + \sigma_{Hr}^2(\pi) \) | Yes |
|                       |                  | \( \hat{r}_{\text{interact}} \) | \( \sigma_r^2(\pi) + \sigma_{Hr}^2(\pi) \) | \( \sigma_r^2(\pi) + \sigma_{Hr}^2(\pi) \) | No |
| Stratification and additional covariates | 1/2 | \( \hat{r}^* \) | \( \sigma_r^2(\pi) + \sigma_{Hr}^2(\pi) \) | \( \sigma_r^2(\pi) + \sigma_{Hr}^2(\pi) \) | No |
| #1/2                  |                  | \( \hat{r}_{\text{adj}}^* \) | \( \sigma_{R^*}^2(\pi) + \sigma_{Hr}^2(\pi) \) | \( \sigma_{R^*}^2(\pi) + \sigma_{Hr}^2(\pi) \) | Yes |
|                       |                  | \( \hat{r}_{\text{interact}}^* \) | \( \sigma_{R^*}^2(\pi) + \sigma_{Hr}^2(\pi) \) | \( \sigma_{R^*}^2(\pi) + \sigma_{Hr}^2(\pi) \) | No |

TABLE 2 Recommendation for regression-based treatment effect estimators

| Regression adjustment | Target allocation | Point estimator | Variance estimator |
|-----------------------|------------------|----------------|--------------------|
| Stratification only   | 1/2              | \( \hat{r}_{\text{adj}} \) | \( \frac{\sigma_r^2(\pi) + \sigma_{Hr}^2(\pi)}{n} \) |
| #1/2                  |                  | \( \hat{r}_{\text{interact}} \) | \( \frac{\sigma_r^2(\pi) + \sigma_{Hr}^2(\pi)}{n} \) |
| Stratification and additional covariates | 1/2 | \( \hat{r}^* \) | \( \frac{\sigma_{R^*}^2(\pi) + \sigma_{Hr}^2(\pi)}{n} \) |
| #1/2                  |                  | \( \hat{r}_{\text{adj}}^* \) | \( \frac{\sigma_{R^*}^2(\pi) + \sigma_{Hr}^2(\pi)}{n} \) |

Before formally stating the asymptotic results, we define several population-level regression (or projection) coefficients. Let \( \Sigma_{RQ} = E[(R - E(R))\{Q - E(Q)\}^\top] \) be the covariance between two random vectors \( R \) and \( Q \), and \( \bar{X} = X - E[X|B(X)] \). We assume that \( X \) has a finite second moment and \( \Sigma_{XX} \) and \( \Sigma_{\bar{X}X} \) are strictly positive-definite. For \( a = 0, 1 \):

\[
\gamma(a) = \arg \min_{r \in R_0} E\{Y_i(a) - E(Y_i(a)) - \{X_i - E(X_i)\}^\top r\}^2 = \Sigma_{XX}^{-1}XY(a),
\]

\[
\beta(a) = \arg \min_{b \in R_0} E\{Y_i(a) - E(Y_i(a)|B_i) - \{X_i - E(X_i|B_i)\}^\top b\}^2 = \Sigma_{XX}^{-1}XY(a),
\]

where \( \gamma(a) \) is the population regression coefficient for regressing \( Y_i(a) \) on \( X_i \) with an intercept, and \( \beta(a) \) is the population regression coefficient for regressing \( \bar{Y}_i(a) = Y_i(a) - E(Y_i(a)|B_i) \) on \( \bar{X}_i = X_i - E(X_i|B_i) \). We also define several sample-level covariates means both overall and stratum-specific: \( \bar{X} = (1/n) \sum_{i=1}^n X_i, \bar{X}_i = (1/n_i) \sum_{i=1}^{n_i} A_i X_i, \bar{X}_0 = (1/n_0) \sum_{i=1}^{n_0} (1 - A_i) X_i, \bar{X}_k = (1/n_{k|0}) \sum_{i=1}^{n_{k|0}} X_i, \bar{X}_{k|1} = (1/n_{k|1}) \sum_{i=1}^{n_{k|1}} A_i X_i, \) and \( \bar{X}_{k|0} = (1/n_{k|0}) \sum_{i=1}^{n_{k|0}} (1 - A_i) X_i \).

We generalize the notations introduced in Section 3 to deal with transformed potential outcomes \( r_i(a), a = 0, 1 \):

\[
\zeta_r^2(\pi) = \frac{1}{\pi} \text{Var}[\hat{r}_i(1)] + \frac{1}{1 - \pi} \text{Var}[\hat{r}_i(0)], \quad \zeta_{Hr}^2 = E[\hat{r}_i(1) - \hat{r}_i(0)]^2,
\]

\[
\zeta_{r*}^2 = \left( \frac{1}{\pi} \hat{r}_i(1) + \frac{1}{1 - \pi} \hat{r}_i(0) \right)^2, \quad \zeta_{Hr*}^2 = \frac{(1 - 2\pi)^2}{\pi^2(1 - \pi)^2} E[q_{B|1} (\hat{r}_i(1) - \hat{r}_i(0))^2],
\]

where \( \hat{r}_i(1) = r_i(a) - E[r_i(a)|B_i] \) and \( \hat{r}_i(0) = E[r_i(a)|B_i] - E[r_i(a)], a = 0, 1 \). By setting \( r_i(a) = Y_i(a) \), the above quantities reduce to those in Section 3. Similarly, we denote their sample analogs as \( \hat{\zeta}_r^2(\pi), \hat{\zeta}_{Hr}^2, \hat{\zeta}_{r*}^2(\pi), \) and \( \hat{\zeta}_{Hr*}^2 \), respectively, and the detailed expressions are given in the Supplementary Material. The explicit forms of several transformed outcomes \( r_i(a) \) are given in subsequent subsections.
4.1 Difference-in-means

First, we consider the regression method that regresses $Y_i$ on $A_i$ and $X_i$ with an intercept but without stratification indicators:

$$Y_i \sim \alpha + A_i \tau + X_i^T \gamma.$$  (4)

Because not adjusting the stratum covariates is contrary to regulatory guidelines and usually decreases efficiency, which becomes more evident shortly, we do not recommend this regression method and only include the main result for completeness. Denote $\hat{\tau}^*$ as the OLS point estimator of $\tau$, and denote $\hat{\tau}$ as the OLS estimator of $\gamma$. Define the transformed outcomes $r_i(a) = Y_i(a) - X_i^T \gamma$, where $\gamma = \pi \gamma(1) + (1 - \pi) \gamma(0)$, and denote the estimates $\hat{r}_i(a) = Y_i(a) - X_i^T \hat{\gamma}$.

**Theorem 1.** Suppose that Assumptions 1–3 hold and $r_i(a) \in R_2$, $a = 0, 1$, then

$$\sqrt{n}(\hat{\tau}^* - \tau) \xrightarrow{d} \mathcal{N}(0, \sigma^2 \pi + \sigma^2 \pi_0 + \sigma^2 \pi_\gamma(\pi)).$$

and the asymptotic variance can be consistently estimated by $\hat{\sigma}^2 \pi(\pi) + \hat{\sigma}^2 \pi_0 + \hat{\sigma}^2 \pi_\gamma(\pi)$.

4.2 Regression adjustment without interaction

The second widely used regression analysis method that adjusts for additional covariates $X_i$ in randomized experiments including clinical trials is to regress $Y_i$ on $A_i$, $I_{i[k]}$ and $X_i$:

$$Y_i \sim \alpha + A_i \tau + \sum_{k=1}^{K-1} \alpha_k I_{i[k]} + X_i^T \beta.$$  (5)

We refer to this regression method as analysis of covariance (ANCOVA). Let $\hat{\tau}^{\text{adj}}$ and $\hat{\sigma}^{2\text{adj}}$ be the OLS point and variance estimators of $\tau$, respectively, and let $\hat{\beta}^{\text{adj}}$ be the OLS estimator of $\beta$. Define the transformed outcomes $r_i^{\text{adj}}(a) = Y_i(a) - X_i^T \hat{\beta}^{\text{adj}}$, where $\beta^{\text{adj}} = \pi \beta(1) + (1 - \pi) \beta(0)$, and denote the estimates $\hat{r}_i^{\text{adj}}(a) = Y_i(a) - X_i^T \hat{\beta}^{\text{adj}}$.

**Theorem 2.** Suppose that Assumptions 1–3 hold and $r_i^{\text{adj}}(a) \in R_2$, $a = 0, 1$, then

$$\sqrt{n}(\hat{\tau}^{\text{adj}} - \tau) \xrightarrow{d} \mathcal{N}(0, \sigma^2 \hat{\tau}^{\text{adj}} + \sigma^2 \hat{\tau}_r^{\text{adj}} + \sigma^2 \pi \hat{\gamma}^{\text{adj}}).$$

and the asymptotic variance can be consistently estimated by $\hat{\sigma}^2 \hat{\tau}^{\text{adj}}(1 - \pi) + \hat{\sigma}^2 \hat{\tau}_r^{\text{adj}} + \hat{\sigma}^2 \pi \hat{\gamma}^{\text{adj}}$. The difference between the asymptotic variances of $\hat{\tau}^*$ and $\hat{\tau}$ is

$$\Delta^{\text{adj}} = -\frac{1}{\pi(1 - \pi)} \beta^{\text{adj}} T \Sigma X \beta^{\text{adj}} + \frac{2(2\pi - 1)}{\pi(1 - \pi)} \beta^{\text{adj}} T \Sigma X \{ \beta(1) - \beta(0) \}.$$

Furthermore, when $\pi = 1/2$, the conclusions hold if Assumption 3 is replaced by Assumption 4.

For equal allocation ($\pi = 1/2$), $\Delta^{\text{adj}} = -4 \beta^{\text{adj}} T \Sigma \Sigma X \Sigma \beta^{\text{adj}} \leq 0$, thus, the asymptotic variance of $\hat{\tau}^{\text{adj}}$ is no greater than that of $\hat{\tau}$. In Corollary 1 we showed that $\hat{\tau}^{\text{adj}}$ and $\hat{\tau}_\text{interact}$ are asymptotically equivalent and they are more efficient than (or at least as efficient as) $\hat{\tau}$. Therefore, $\hat{\tau}^{\text{adj}}$ is more efficient than (or at least as efficient as) $\hat{\tau}$, $\hat{\tau}^{\text{adj}}$, and $\hat{\tau}_\text{interact}$. We show later that in general, $\hat{\tau}^{\text{adj}}$ is also more efficient than $\hat{\tau}^*$ (see Corollary 2 and Theorem 4). Moreover, as $\hat{\sigma}^2 \hat{\tau}_r^{\text{adj}} = 0$ and $\hat{\sigma}^2 \hat{\tau}^{\text{adj}}(1 - \pi) = \hat{\sigma}^2 \hat{\tau}^{\text{adj}}(\pi)$, the usual OLS variance estimator $\hat{\sigma}^{2\text{adj}}$ is consistent.

However, for unequal allocation ($\pi \neq 1/2$), $\hat{\tau}^{\text{adj}}$ may hurt the precision when compared with that of $\hat{\tau}_\text{adj}$, and $\hat{\sigma}^{2\text{adj}}$ can be anti-conservative. The same phenomenon has been observed when estimating the sample average treatment effect in completely randomized experiments under finite population framework.\(^3\)
Remark 2. Under stronger assumptions than those in Theorem 2, Wang et al.\textsuperscript{17} also established the asymptotic property of the ANCOVA estimator \( \hat{\tau}_{adj} \) by treating it as an M-estimator. They required that strong balance should be achieved for stratified randomization, that is, \( q_{[k]} = 0 \), in Assumption 3. Therefore, their results do not apply to, for example, Wei’s urn design or minimization. Theorem 2 relaxes this requirement by allowing \( q_{[k]} \geq 0 \) in Assumption 3 in general, and it emphasizes that Assumption 3 can be weakened to Assumption 4 under equal allocation, thus covering a broader range of covariate-adaptive randomization methods.

### 4.3 Regression adjustment with interaction

In this section, we consider a regression that follows the same principle as that described in Section 3.3 while adding additional covariates \( X_i \):

\[
Y_i \sim \alpha + A_i \tau + \sum_{k=1}^{K-1} \alpha_k I_{[i \in [k]]} + \sum_{k=1}^{K-1} \nu_k A_i (I_{[i \in [k]]} - p_{n[k]}) + X_i^T \beta + A_i (X_i - \bar{X})^T \xi.
\]

(6)

Denote \( \hat{\tau}_{interact}^* \) and \( \hat{\sigma}_{interact}^{2*} \) as the OLS point and variance estimators of \( \tau \), respectively. Similar to \( \hat{\tau}_{interact} \), we center the stratification indicators and covariates at their sample means in the interactions so that \( \hat{\tau}_{interact}^* \) alone can be used to estimate the treatment effect \( \tau \). As shown in our proof in the Supplementary Material, regression (6) is equivalent to running two regressions. First, we separately regress the outcomes \( Y_i \) on the stratification indicators \( I_{[i \in [k]]} \) and covariates \( X_i \) in the treatment and control groups, and obtain the OLS estimators \( \hat{\beta}(1) \) and \( \hat{\beta}(0) \) of the coefficients of \( X_i \). Then:

\[
\hat{\tau}_{interact}^* = \sum_{k=1}^{K} p_{n[k]} \left[ \{ \bar{Y}_{[k]1} - (\bar{X}_{[k]1} - \bar{X}_{[k]})^T \hat{\beta}(1) \} - \{ \bar{Y}_{[k]0} - (\bar{X}_{[k]0} - \bar{X}_{[k]})^T \hat{\beta}(0) \} \right].
\]

Remark 3. By comparing the above expression to (3), \( \hat{\tau}_{interact}^* \) can be viewed as augmenting \( \hat{\tau}_{interact} \) by further adjusting for the additional covariates \( X_i \) when estimating the stratum-specific treatment effect. This covariate adjustment has also been used for estimators under simple or stratified block randomization.\textsuperscript{19,32}

Define the transformed outcomes \( r_{i,interact}(a) = Y_i(a) - X_i^T \beta_{interact} \) and the estimates \( \hat{r}_{i,interact}(a) = Y_i(a) - X_i^T \hat{\beta}_{interact} \), where \( \beta_{interact} = (1 - \pi) \beta(1) + \pi \beta(0) \) and \( \hat{\beta}_{interact} = (1 - \pi) \hat{\beta}(1) + \pi \hat{\beta}(0) \).

**Theorem 3.** Suppose that Assumptions 1, 2, and 4 hold and \( r_{i,interact}(a) \in R_2, \ a = 0, 1 \), then

\[
\sqrt{n}(\hat{\tau}_{interact}^* - \tau) \xrightarrow{d} \mathcal{N}(0, \xi_{2\tau_{interact}}^2(\pi) + \xi_{H\tau_{interact}}^2),
\]

\[
n\hat{\sigma}_{interact}^{2*} \xrightarrow{p} \xi_{2\tau_{interact}}^2(1 - \pi) = \left\{ \frac{1}{\pi(1 - \pi)} - 3 \right\} \{ \beta(1) - \beta(0) \}^T \Sigma \bar{X} \{ \beta(1) - \beta(0) \},
\]

and the asymptotic variance can be consistently estimated by \( \xi_{2\tau_{interact}}^2(\pi) + \xi_{H\tau_{interact}}^2 \). Furthermore, the difference between the asymptotic variances of \( \hat{\tau}_{interact}^* \) and \( \hat{\tau}_{interact} \) is

\[
\Delta_{interact-interact} = \frac{1}{\pi(1 - \pi)} \beta_{interact}^T \Sigma \bar{X} \beta_{interact} \leq 0,
\]

and the difference between the asymptotic variances of \( \hat{\tau}_{interact}^* \) and \( \hat{\tau}_{adj}^* \) is

\[
\Delta_{interact-adj} = \frac{1}{\pi(1 - \pi)} (\beta_{interact} - \beta_{adj})^T \Sigma \bar{X} (\beta_{interact} - \beta_{adj})^T - \xi_{2\tau_{adj}}^2 \leq 0.
\]

Remark 4. The asymptotic normality of \( \hat{\tau}_{interact}^* \) and the consistency of the non-parametric variance estimator, with slight modification, still hold even if the target treatment proportion differs across strata. The details are given in the Supplementary Material.
Theorem 3 implies that the regression with treatment-by-covariate interactions results in an estimator $\hat{\tau}_{\text{interact}}^*$ whose asymptotic variance is no greater than that of $\hat{\tau}_{\text{interact}}$, the most efficient estimator among $\hat{\tau}$, $\hat{\tau}_{\text{adj}}$, and $\hat{\tau}_{\text{interact}}$. That is, using additional covariates in an interaction regression will improve, or at least not hurt, the efficiency. As $\Delta_{\text{interact} - \text{adj}}^* \leq 0$, the regression with treatment-by-covariate interactions is generally more efficient than ANCOVA. The higher efficiency of $\hat{\tau}_{\text{interact}}^*$ as compared with that of $\hat{\tau}^*$ is later confirmed by Theorem 4. As $1/\{\pi(1 - \pi)\} - 3 \geq 1$, the OLS variance estimator $\hat{\tau}_{\text{interact}}^*$ may underestimate the asymptotic variance even when $\pi = 1/2$ and thus should not be used. Theorem 3 provides a consistent variance estimator that can be used to construct a valid confidence interval or test for $\tau$. We note that the asymptotic variance of $\hat{\tau}_{\text{interact}}^*$ and the proposed variance estimator are invariant with respect to randomization methods.

**Corollary 2.** When $\pi = 1/2$, suppose that Assumptions 1, 2, and 4 hold, $r_{\text{adj}}(a) \in R_2$ and $r_{\text{interact}}(a) \in R_2$, $a = 0, 1$, then $\hat{\tau}_{\text{adj}}^*$ and $\hat{\tau}_{\text{interact}}^*$ are asymptotically equivalent and their asymptotic variances are smaller than or equal to that of $\hat{\tau}_{\text{interact}}$. Furthermore, both $n\sigma_{\text{adj}}^2$ and $\zeta_{\text{interact}}^2$ are consistent variance estimators.

Corollary 2 implies that, when $\pi = 1/2$, $\hat{\tau}_{\text{adj}}^*$ is as efficient as $\hat{\tau}_{\text{interact}}^*$, neither of which hurts the efficiency, as compared with $\hat{\tau}$, $\hat{\tau}_{\text{adj}}$, and $\hat{\tau}_{\text{interact}}$, and it is valid and efficient to use the OLS point and variance estimators in the ANCOVA, that is, regression (5), to estimate and infer $\tau$. For equal allocation, we recommend the ANCOVA for two reasons: first, it has significantly fewer parameters than the regression with treatment-by-covariate interactions, and second, it is easily implemented by standard statistical packages.

### 4.4 Summary and recommendation

At the end of this section, we present a summary of the asymptotic variances and variance estimators in Table 1, which provides answers to the open question raised by Wang et al.\(^2\) (stated in the introduction). Based on the above arguments, when taking into account both simplicity and efficiency, Table 2 shows our recommendations for regression estimators that adjust for additional covariates $X_0$. As in Section 3, these recommendations apply to most of the widely used covariate-adaptive randomization methods, including stratified randomization and minimization.

### 5 S-OPTIMAL

In this section, we will show that $\hat{\tau}_{\text{interact}}^*$ is optimal among the following class of estimators:

$$S = \{ \hat{\tau}^*(\hat{\eta}(1), \hat{\eta}(0)), \hat{\tau}_{\text{adj}}^*(\hat{\eta}(1), \hat{\eta}(0)), \hat{\tau}_{\text{interact}}^*(\hat{\eta}(1), \hat{\eta}(0)) \},$$

where $\hat{\eta}(1)$ and $\hat{\eta}(0)$ are regression-adjusted vectors (in fact, they can be any estimators as long as they have limits, $\eta(1)$ and $\eta(0)$, in probability) and

$$\hat{\tau}^*(\hat{\eta}(1), \hat{\eta}(0)) = \{ \bar{Y}_1 - (\bar{X}_1 - \bar{X})^T \hat{\eta}(1) \} - \{ \bar{Y}_0 - (\bar{X}_0 - \bar{X})^T \hat{\eta}(0) \},$$

$$\hat{\tau}_{\text{adj}}^*(\hat{\eta}(1), \hat{\eta}(0)) = \sum_{k=1}^K \omega_{k[1]} \left[ \{ \bar{Y}_{[k]} - (\bar{X}_{[k]} - \bar{X})^T \hat{\eta}(1) \} - \{ \bar{Y}_{[k]} - (\bar{X}_{[k]} - \bar{X})^T \hat{\eta}(0) \} \right],$$

$$\hat{\tau}_{\text{interact}}^*(\hat{\eta}(1), \hat{\eta}(0)) = \sum_{k=1}^K p_{[k]} \left[ \{ \bar{Y}_{[k]} - (\bar{X}_{[k]} - \bar{X})^T \hat{\eta}(1) \} - \{ \bar{Y}_{[k]} - (\bar{X}_{[k]} - \bar{X})^T \hat{\eta}(0) \} \right].$$

**Definition 1.** Given the randomization mechanism, an estimator in $S$ is $S$-optimal if it has the smallest asymptotic variance among the estimators in $S$.

The definition of the $S$-optimal estimator appeared in Li and Ding for completely randomized experiments and rerandomization, which we extend to covariate-adaptive randomization. In additional to $\hat{\tau}^*$, $\hat{\tau}_{\text{adj}}^*$, and $\hat{\tau}_{\text{interact}}^*$, $S$ contains other regression estimators, such as those with added interactions $A_i(X_i - \bar{X})$ in regression (4) or deleted interactions $A_i(X_i - \bar{X})$ from regression (6). Our next theorem shows that $\hat{\tau}_{\text{interact}}^*$ is $S$-optimal, so those regressions need not be investigated. Define $r_{\text{gen}}(a) = Y_i(a) - X_i^T \{ (1 - \pi)\eta(1) + \pi\eta(0) \}$, $a = 0, 1$.
Theorem 4. Suppose that Assumptions 1–3 hold, \( r_{\text{gen}}(a) \in R_2 \) and \( \hat{n}(a) \xrightarrow{P} \eta(a) \), \( a = 0, 1 \), then \( \hat{\tau}^*_{\text{interact}} \) is S-optimal.

Remark 5. We note that \( \hat{\tau}^*_{\text{interact}} \) is not optimal among all regression adjusted estimators for at least two reasons. First, there exists weights \( \hat{o}_{k|j} \) such that \( \sum_{k=1}^{K} \hat{o}_{k|j} [ \{ \hat{Y}_{k|j} - \hat{X}_{k|j} \}^T \hat{\beta}(1) - \{ \hat{Y}_{k|j} - \hat{X}_{k|j} \}^T \hat{\beta}(0) ] \) has a smaller asymptotic variance than \( \hat{\tau}^*_{\text{interact}} \). However, these weights \( \hat{o}_{k|j} \) depend on the unknown potential outcomes in a complicated form. We leave to future work the investigation of semi-parametric efficiency under covariate-adaptive randomization. Second, to further improve efficiency, one can use the stratum-specific adjusted vectors \( \hat{\beta}_{k|j}(1) \) and \( \hat{\beta}_{k|j}(0) \) within stratum \( k \) instead of the common adjusted vectors \( \hat{\beta}(1) \) and \( \hat{\beta}(0) \) for all strata,

6 | SIMULATION STUDY

In this section, we report the results of a simulation study in which the empirical performances of the six proposed estimators were examined with respect to treatment effect estimation and asymptotic variance. For comparison, the OLS and Huber-White variance estimators are also considered. For \( a \in \{0, 1\} \) and \( 1 \leq i \leq n \), the potential outcomes are generated using the equation:

\[
Y_i(a) = \mu_a + g_a(X_i) + \sigma_a(X_i) \varepsilon_i(a),
\]

where \( X_i, g_a(X_i), \) and \( \sigma_a(X_i) \) are specified below for three different models. In each of the models, \( (X_i, \varepsilon_i(0), \varepsilon_i(1)), 1 \leq i \leq n \), are i.i.d, and both \( \varepsilon_i(0) \) and \( \varepsilon_i(1) \) follow the standard normal distribution. The number of units \( n \) is 1000.

Model 1: \( X_i \) is a five-dimensional vector,

\[
g_0(X_i) = g_1(X_i) = \sum_{j=1}^{5} \beta_j X_{ij},
\]

where \( X_{i1} \sim \text{Beta}(2, 2), X_{i2} \) takes values in \( \{1, 2, 3, 4\} \) with equal probability, \( X_{i3} \sim \text{Unif}[−2, 2], X_{i4} \) takes values in \( \{1, 2, 3\} \) with respective probabilities of \( 0.3, 0.6, 0.1 \), \( X_{i5} \sim \mathcal{N}(0, 1) \), and they are all independent of each other. \( \sigma_0(X_i) = 1, \ \sigma_1(X_i) = 3, \ \beta = (2, 8, 10, 3, 6)^T \). \( X_{i2} \) and \( X_{i4} \) are used for randomization, and \( X_{i1} \) and \( X_{i3} \) are used as the additional covariates.

Model 2: \( X_i \) is a four-dimensional vector,

\[
g_0(X_i) = a_1 X_{i1} + \log(\alpha_3 X_{i1} \log(X_{i3} + 1) + 1) + \alpha_4 e^{X_{i3}},
\]

\[
g_1(X_i) = \alpha_2 X_{i2}^2 + \log(\alpha_3 X_{i1} \log(X_{i3} + 1) + 1),
\]

where \( X_{i1} \sim \text{Gamma}(2, 1), X_{i2} \) takes values in \( \{1, 2, 3\} \) with respective probabilities of \( 0.3, 0.6, 0.1 \), \( X_{i3} \sim \text{Poisson}(3), X_{i4} \sim \text{Beta}(2, 2) \), and they are all independent of each other. \( \sigma_0(X_i) = 2, \ \sigma_1(X_i) = 1, \ \alpha = (5, 8, 3, 12)^T \). \( X_{i15} \) is the stratified variable of \( X_{i1} \), determined by its relative value of \( 2.5 \). \( X_{i15} \) and \( X_{i2} \) are used for randomization, and \( X_{i1} \) and \( X_{i3} \) are used as the additional covariates.

Model 3: \( X_i \) is a four-dimensional vector,

\[
g_0(X_i) = \sum_{j=1}^{4} \beta_j X_{ij}, \quad g_1(X_i) = \beta_1 \log(X_{i1})X_{i4},
\]

where \( X_{i1} \sim \text{Beta}(3, 4), X_{i2} \sim \text{Unif}[−2, 2], X_{i3} = X_{i1}X_{i2}, X_{i4} \) takes values in \( \{3, 5\} \) with respective probabilities of \( 0.6, 0.4 \). \( X_{i1}, X_{i2}, X_{i4} \) are independent of each other. \( \sigma_0(X_i) = X_{i35}, \ \sigma_1(X_i) = 2X_{i25}, \ \beta = (20, 7, 5, 6)^T \), where \( X_{i25} \) is the stratified variable of \( X_{i2} \), if \( X_{i2} > 1, X_{i25} = 2, \) and \( X_{i25} = 1 \) otherwise, and \( X_{i35} \) is the stratified variable of \( X_{i3} \), if \( X_{i3} > 0, X_{i35} = 2, \) and \( X_{i35} = 1 \) otherwise. \( X_{i25} \) and \( X_{i4} \) are used for randomization, and \( X_{i1} \) and \( X_{i3} \) are used as the additional covariates.

Here we present the simulation results of the six estimators under simple randomization, the stratified block randomization, and minimization in Table 3 (equal allocation) and Table 4 (unequal allocation). The block size used in the
stratified block randomization was 6. The biased-coin probability and the weight used in minimization were 0.75 and (0.5, 0.5)², respectively. The bias, standard deviation (SD), and root mean squared error (RMSE) of the treatment effect estimators, standard error (SE) estimators, and empirical coverage probabilities (CP) of 95% confidence intervals were computed using 10³ replications. Note that, under minimization, the SE estimators of $\hat{\tau}_{adj}$ and $\hat{\tau}^{*}_{adj}$ are calculated according to the known theory, but they are not available for other estimators. Additional simulation results under other three covariate-adaptive designs can be found in the Supplementary Material.

Table 3 presents the simulation results under equal allocation. First, the six treatment effect estimators all have small finite-sample biases, and their root mean squared errors are almost identical to their SDs. Under simple randomization, the SDs of $\hat{\tau}_{adj}$ and $\hat{\tau}_{interact}$ are comparable and smaller than that of $\hat{\tau}$. Regarding stratified block randomization, which achieves strong balance, the SDs of these three estimators are nearly the same, which is consistent with the theoretical results. Under minimization, the first three estimators also tend to have similar SDs. Moreover, by adding the stratification indicators into the regression, almost identical SDs are obtained by the different randomization methods, regardless of whether the interaction terms are included. Besides, making use of additional covariates can further improve the precision when estimating the treatment effect. The relationship among the last three estimators is the same as that among the first three estimators.

Table 3 also confirms that the proposed variance estimators coincide with the simulated variance of the treatment effect estimators and lead to the expected 95% coverage probability. Moreover, the OLS and Huber-White variance estimators are valid for $\hat{\tau}_{adj}$ and $\hat{\tau}^{*}_{adj}$, so in this situation, we can rely on the output of standard statistical packages. For other scenarios, however, the OLS and Huber-White variance estimators are generally not valid and can be either larger or smaller than the true values. In particular, both the variance estimators are anti-conservative for $\hat{\tau}_{interact}$, which concurs with the results of Bugni et al.¹⁶

Table 4 shows the simulation results under unequal allocation ($\pi = 2/3$). First, the biases of the six estimators are negligible, and the root mean squared errors are almost identical to the SDs. Second, unlike equal allocation, adjustment of the stratification indicators does not guarantee a gain in efficiency when the stratification-by-treatment interactions are not included; the SD of $\hat{\tau}_{adj}$ is larger than that of $\hat{\tau}$ under simple randomization for models 2 and 3. However, adding the interactions always reduces the SD. Similar conclusions can be drawn for the last three estimators. Third, adjusting for additional covariates can further improve the efficiency. Among the six estimators, $\hat{\tau}^{*}_{interact}$ is the most efficient, and its variance is invariant under different randomization methods, which reflects the benefit of adding treatment-by-covariate interactions under unequal allocation. Finally, the OLS and Huber-White variance estimators of $\hat{\tau}_{interact}$ or $\hat{\tau}^{*}_{interact}$ are no longer similar, and they are not valid under most of the considered scenarios. The proposed variance estimators, in contrast, are still valid and lead to the expected 95% coverage probability under all scenarios.

## 7 | CLINICAL TRIAL EXAMPLE

In this section, we consider a clinical trial example based on the synthetic data of the Nefazodone CBASP trial, which was conducted to compare the efficacy of three treatments for chronic depression.¹⁵ We focus on two of the treatments, Nefazodone and the combination of Nefazodone and the cognitive behavioral-analysis system of psychotherapy (CBASP). To generate the synthetic data, we first fit a non-parametric spline using the function bigss in the R package bigspline with six selected covariates: AGE, HAMD17, HAMD24, HAMD_COGIN, Mstatus2, and TreatPD, which are detailed in Table 5.

Then, we implement simple randomization and stratified block randomization to obtain the treatment assignments for both the equal ($\pi = 1/2$) and unequal ($\pi = 2/3$) allocations. Among the six covariates, we use the stratified HAMD17 and Mstatus2 in stratified block randomization, where the stratified HAMD17 is determined by the relative values of 18 and 21. Once the treatment assignments are produced, we generate patient outcomes through the fitted model. For data analysis, AGE, HAMD17, HAMD24, HAMD_COGIN, and TreatPD are used as the additional covariates. The analysis was performed according to the recommendations given in Table 2, and the results are shown in Table 6.

As seen in Table 6, for simple randomization, adjusting for stratification and additional baseline covariates clearly improves the estimation precision. Compared with $\hat{\tau}$, the variance reductions for $\hat{\tau}_{adj}$ and $\hat{\tau}^{*}_{adj}$ are about 9% and 15%, respectively, under equal allocation, and the reductions for $\hat{\tau}_{interact}$ and $\hat{\tau}^{*}_{interact}$ are more significant under unequal allocation, that is, approximately 13% and 24%, respectively. With respect to stratified block randomization, which achieves strong balance, the variance reduction of $\hat{\tau}_{adj}$ under equal allocation is close to 0 and that of $\hat{\tau}_{interact}$ under unequal allocation is exactly 0, as expected. Moreover, the variance reductions of $\hat{\tau}^{*}_{adj}$ and $\hat{\tau}^{*}_{interact}$ are moderate, approximately 4.5% and
## Table 3

Simulated biases, SDs, SEs, and coverage probabilities for different estimators and randomization methods under equal allocation ($\pi = 1/2$)

| Mod. Est. | Simple randomization | Stratified block randomization | Minimization |
|-----------|----------------------|--------------------------------|-------------|
|           | SE                   | CP                             | SE          | CP          | SE          | CP          |
|           | Bias     | SD         | RMSE    | NP | OLS | HW | Bias     | SD         | RMSE    | NP | OLS | HW |
| 1         | $\hat{\tau}$     | -0.02     | 1.03    | 1.03 | 1.01 | 1.01 | 0.94     | 0.94     | 0.94    | 0.05 | 0.84 | 0.84 |
|           |          |            |         | NP | OLS | HW |          |           |         |       |     |     |
|           | $\hat{\tau}_\text{adj}$ | -0.03   | 0.87    | 0.87 | 0.83 | 0.83 | 0.94     | 0.94     | 0.94    | 0.04 | 0.84 | 0.84 |
|           |          |            |         | NP | OLS | HW |          |           |         |       |     |     |
|           | $\hat{\tau}_\text{interact}$ | -0.03  | 0.87    | 0.87 | 0.83 | 0.83 | 0.95     | 0.95     | 0.95    | 0.04 | 0.84 | 0.84 |
|           |          |            |         | NP | OLS | HW |          |           |         |       |     |     |
| 2         | $\hat{\tau}$     | -0.02     | 0.88    | 0.88 | 0.90 | 0.90 | 0.95     | 0.95     | 0.95    | 0.01 | 0.68 | 0.68 |
|           |          |            |         | NP | OLS | HW |          |           |         |       |     |     |
|           | $\hat{\tau}_\text{adj}$ | -0.04   | 0.66    | 0.66 | 0.67 | 0.68 | 0.95     | 0.95     | 0.95    | 0.01 | 0.67 | 0.67 |
|           |          |            |         | NP | OLS | HW |          |           |         |       |     |     |
|           | $\hat{\tau}_\text{interact}$ | -0.00  | 0.65    | 0.65 | 0.67 | 0.31 | 0.31   | 0.95     | 0.95    | 0.95    | 0.01 | 0.67 | 0.67 |
|           |          |            |         | NP | OLS | HW |          |           |         |       |     |     |
| 3         | $\hat{\tau}$     | 0.03      | 1.99    | 1.99 | 1.99 | 2.00 | 2.00    | 0.95     | 0.95    | 0.95   | -0.09 | 1.96 | 1.97 |
|           |          |            |         | NP | OLS | HW |          |           |         |       |     |     |
|           | $\hat{\tau}_\text{adj}$ | 0.08   | 1.95    | 1.95 | 1.93 | 1.94 | 1.94   | 0.94     | 0.94    | 0.94   | -0.09 | 1.96 | 1.97 |
|           |          |            |         | NP | OLS | HW |          |           |         |       |     |     |
|           | $\hat{\tau}_\text{interact}$ | 0.04  | 1.95    | 1.95 | 1.93 | 1.77 | 1.76   | 0.94     | 0.92    | 0.92   | -0.09 | 1.96 | 1.97 |
|           |          |            |         | NP | OLS | HW |          |           |         |       |     |     |
|           | $\hat{\tau}_\text{interact}$ | -0.03 | 1.51    | 1.51 | 1.54 | 1.54 | 0.95   | 0.95    | 0.95   | -0.05 | 1.52 | 1.52 |
|           |          |            |         | NP | OLS | HW |          |           |         |       |     |     |
|           | $\hat{\tau}_\text{interact}$ | 0.02  | 1.47    | 1.47 | 1.48 | 1.49 | 1.49   | 0.95     | 0.95    | 0.95   | -0.06 | 1.51 | 1.51 |
|           |          |            |         | NP | OLS | HW |          |           |         |       |     |     |
|           | $\hat{\tau}_\text{interact}$ | 0.01  | 1.46    | 1.46 | 1.48 | 0.70 | 0.69   | 0.95     | 0.65    | 0.65   | -0.02 | 1.51 | 1.51 |

Abbreviations: CP, coverage probability; Est., estimator; HW, the Huber-White variance estimator; Mod., model; NP, the proposed non-parametric variance estimator; OLS, the ordinary least-squares variance estimator; RMSE, root mean squared error; SD, standard deviation; SE, standard error.
### Table 4: Simulated biases, SDs, SEs, and coverage probabilities for different estimators and randomization methods under unequal allocation ($\pi = 2/3$)

| Mod. Est. | Simple randomization | Stratified block randomization | Minimization |
|-----------|----------------------|-------------------------------|--------------|
|           | SE                   | CP               | SE | CP | SE | CP | SE | CP | SE | CP | SE | CP | SE | CP | SE | CP |
|           | Bias | SD | RMSE | NP | OLS | HW | NP | OLS | HW | Bias | SD | RMSE | NP | OLS | HW | Bias | SD | RMSE | NP | OLS | HW |
| 1         | \( \hat{\tau} \) | 0.00 | 1.08 | 1.08 | 1.07 | 1.07 | 0.95 | 0.96 | 0.95 | -0.03 | 0.89 | 0.89 | 0.97 | 0.98 | -0.06 | 0.90 | 0.91 | -   | 1.08 | 1.07 | -   | 0.98 | 0.98 |
|           | \( \hat{\tau}_{\text{adj}} \) | 0.00 | 0.88 | 0.88 | 0.88 | 0.90 | 0.88 | 0.95 | 0.95 | -0.03 | 0.89 | 0.89 | 0.94 | 0.94 | -0.05 | 0.90 | 0.90 | -   | 0.89 | 0.88 | -   | 0.94 | 0.94 |
|           | \( \hat{\tau}_{\text{interact}} \) | 0.00 | 0.88 | 0.88 | 0.87 | 0.90 | 0.88 | 0.95 | 0.95 | -0.03 | 0.89 | 0.87 | 0.94 | 0.94 | -0.04 | 0.90 | 0.90 | 0.87 | 0.89 | 0.87 | 0.94 | 0.95 | 0.94 |
| 2         | \( \hat{\tau} \) | 0.05 | 0.85 | 0.85 | 0.86 | 1.05 | 0.86 | 0.95 | 0.99 | 0.95 | 0.01 | 0.73 | 0.73 | 0.70 | 1.05 | 0.86 | 0.94 | 0.99 | 0.98 | -0.07 | 0.72 | 0.72 | -   | 1.04 | 0.86 | 1.00 | 0.98 |
|           | \( \hat{\tau}_{\text{adj}} \) | -0.04 | 0.88 | 0.88 | 0.82 | 0.65 | 0.82 | 0.93 | 0.86 | 0.93 | 0.01 | 0.73 | 0.73 | 0.70 | 0.66 | 0.82 | 0.94 | 0.92 | 0.97 | -0.02 | 0.72 | 0.72 | -   | 0.66 | 0.81 | 0.93 | 0.98 |
|           | \( \hat{\tau}_{\text{interact}} \) | 0.01 | 0.73 | 0.73 | 0.70 | 0.27 | 0.37 | 0.94 | 0.55 | 0.69 | 0.01 | 0.73 | 0.73 | 0.70 | 0.27 | 0.37 | 0.94 | 0.54 | 0.68 | -0.02 | 0.71 | 0.71 | 0.70 | 0.27 | 0.37 | 0.95 | 0.54 | 0.68 |
| 3         | \( \hat{\tau} \) | 0.11 | 1.72 | 1.72 | 1.80 | 2.40 | 1.80 | 0.96 | 0.99 | 0.96 | 0.05 | 1.80 | 1.80 | 1.76 | 2.40 | 1.80 | 0.94 | 0.99 | 0.95 | -0.15 | 1.78 | 1.78 | -   | 2.40 | 1.80 | -   | 0.99 | 0.95 |
|           | \( \hat{\tau}_{\text{adj}} \) | 0.10 | 1.79 | 1.80 | 1.85 | 2.30 | 1.86 | 0.96 | 0.98 | 0.96 | 0.05 | 1.80 | 1.80 | 1.76 | 2.29 | 1.85 | 0.94 | 0.99 | 0.95 | -0.13 | 1.78 | 1.78 | -   | 2.29 | 1.85 | -   | 0.99 | 0.96 |
|           | \( \hat{\tau}_{\text{interact}} \) | 0.09 | 1.69 | 1.69 | 1.76 | 2.15 | 1.56 | 0.96 | 0.98 | 0.94 | 0.05 | 1.80 | 1.80 | 1.76 | 2.14 | 1.55 | 0.94 | 0.97 | 0.90 | -0.12 | 1.77 | 1.78 | 1.76 | 2.14 | 1.56 | 0.94 | 0.98 | 0.91 |
|           | \( \hat{\tau}^* \) | 0.04 | 1.72 | 1.61 | 1.64 | 1.72 | 1.65 | 0.96 | 0.97 | 0.96 | -0.07 | 1.68 | 1.68 | 1.62 | 1.72 | 1.64 | 0.95 | 0.96 | 0.95 | -0.14 | 1.63 | 1.64 | 1.72 | 1.64 | 0.96 | 0.96 | 0.96 |
|           | \( \hat{\tau}_{\text{adj}}^* \) | 0.04 | 1.72 | 1.72 | 1.72 | 1.57 | 1.73 | 0.96 | 0.93 | 0.96 | -0.07 | 1.68 | 1.68 | 1.62 | 1.57 | 1.72 | 0.95 | 0.94 | 0.96 | -0.13 | 1.64 | 1.65 | 1.57 | 1.72 | 0.93 | 0.95 |
|           | \( \hat{\tau}^*_\text{interact} \) | 0.08 | 1.42 | 1.42 | 1.45 | 0.85 | 0.62 | 0.95 | 0.76 | 0.61 | -0.02 | 1.50 | 1.49 | 1.45 | 0.85 | 0.62 | 0.94 | 0.73 | 0.57 | -0.12 | 1.47 | 1.47 | 1.45 | 0.85 | 0.62 | 0.94 | 0.76 | 0.62 |

Abbreviations: CP, coverage probability; Est., estimator; HW, the Huber-White variance estimator; Mod., model; NP, the proposed non-parametric variance estimator; OLS, the ordinary-least-squares variance estimator; RMSE, root mean squared error; SD, standard deviation; SE, standard error.
TABLE 5 Description of selected covariates

| Variable       | Description                                      |
|----------------|--------------------------------------------------|
| AGE            | Age of patients in years                         |
| HAMD17         | Total HAMD-17 score                              |
| HAMD24         | Total HAMD-24 score                              |
| HAMD_COGIN     | HAMD cognitive disturbance score                 |
| Mstatus2       | Marriage status: 1 if married or living with someone and 0 otherwise |
| TreatPD        | Treated past depression: 1 yes and 0 no          |

Abbreviation: HAMD, Hamilton Rating Scale for Depression.

TABLE 6 Estimates, 95% confidence intervals, and variance reductions under simple randomization and stratified block randomization for synthetic Nefazodone CBASP trial data

| Randomization method         | Equal allocation (π = 1/2) | Variance reduction | Unequal allocation (π = 2/3) | Variance reduction |
|------------------------------|----------------------------|--------------------|------------------------------|--------------------|
|                              | Estimator     | Estimate  | 95% CI |                     | Estimator     | Estimate  | 95% CI |                     |                   |
| Simple randomization         | \( \hat{\tau} \)  | -4.70     | (-5.20, -4.19) | -  | \( \hat{\tau} \)  | -4.69     | (-5.25, -4.12) | -  |                  |
|                              | \( \hat{\tau}_{adj} \)  | -4.69     | (-5.17, -4.21) | 9.43% | \( \hat{\tau}_{interact} \)  | -4.69     | (-5.22, -4.16) | 12.90% |                  |
|                              | \( \hat{\tau}^*_{adj} \)  | -4.69     | (-5.16, -4.23) | 14.85% | \( \hat{\tau}^*_{interact} \)  | -4.69     | (-5.19, -4.20) | 23.68% |                  |
| Stratified block randomization | \( \hat{\tau} \)  | -4.68     | (-5.15, -4.21) | -  | \( \hat{\tau} \)  | -4.69     | (-5.22, -4.17) | -  |                  |
|                              | \( \hat{\tau}_{adj} \)  | -4.68     | (-5.16, -4.21) | -1.64% | \( \hat{\tau}_{interact} \)  | -4.69     | (-5.22, -4.17) | 0%     |                  |
|                              | \( \hat{\tau}^*_{adj} \)  | -4.68     | (-5.15, -4.22) | 4.55% | \( \hat{\tau}^*_{interact} \)  | -4.70     | (-5.20, -4.21) | 12.38% |                  |

Abbreviation: CI, confidence interval.

12% under equal allocation and unequal allocation, respectively. Finally, the performance of each of the recommended estimators is comparable for different randomization methods.

8 | DISCUSSION

Linear regression is widely used to analyze randomized clinical trials and other experiments with the hope of improving efficiency. In this article, we investigated the theoretical properties of six widely used regression-based treatment effect estimators under covariate-adaptive randomization, without imposing any modeling assumptions on the potential outcomes and covariates. We showed that these estimators are all consistent and asymptotically normal, albeit with possibly different efficiency. We provided non-parametric variance estimators to construct a valid confidence interval or test for the treatment effect and discussed the conditions under which the usual OLS variance estimator is valid.

Taking into account both simplicity and efficiency, our final recommendations to practitioners are as follows. For equal allocation, we recommend using regression (without interactions), which adjusts for the stratification indicators and additional covariates if available, and the use of the usual OLS variance estimator. For unequal allocation, we recommend using regression with interactions and the proposed non-parametric variance estimator.

This article focused on using linear regression to estimate and infer the treatment effect under covariate-adaptive randomization with two treatments. It would be interesting to extend the results to more complicated settings, such as linear regression analysis for multiple treatments, for binary outcomes using logistic regression, and when there are missing values. Moreover, in this article, we assumed that the number of strata and the number of covariates were fixed. It would be of interest to generalize the results to high-dimensional settings. Finally, although our results provide insights for using more general classes of models, theoretical results have yet to be established, especially regarding semi-parametric efficiency, which merits further investigation.

Last, it is worth mentioning that the assumptions in this article, although reasonable and commonly used in the literature, may still be violated in practice. First, it is critical to understand and quantify the impact of different types of biases.
that can be present in randomized clinical trials, including chronological bias, selection bias, and accidental bias,\textsuperscript{8,36} which is beyond the scope of this article and left for future work. For unequal allocation, the randomization procedures that preserve the allocation ratio at each allocation step may protect against some of these biases (eg, selection bias and accidental bias), and the procedures without such allocation ratio preserving property may be problematic with the rerandomization test.\textsuperscript{37} This aspect also requires further investigation. Moreover, in practice some important covariates may have missing values or cannot be observed,\textsuperscript{17,38,39} and their impact should be evaluated. In addition, more complicated randomization schemes could be considered, such as those balancing for continuous covariates\textsuperscript{40,41} and those that also incorporate patient outcomes into the randomization procedure.\textsuperscript{42}

**AUTHOR CONTRIBUTIONS**
Wei Ma, Fuyi Tu, and Hanzhong Liu designed research, performed research, analyzed data, and wrote the article.

**ACKNOWLEDGEMENTS**
Dr. Wei Ma was supported by the National Key R&D Program of China (Grant No. 2018YFC2000302) and the National Natural Science Foundation of China (Grant No. 12171476). Dr. Hanzhong Liu was supported by a Grant 12071242 from the National Natural Science Foundation of China and the Guo Qiang Institute of Tsinghua University.

**CONFLICT OF INTEREST**
The authors declare no potential conflict of interests.

**DATA AVAILABILITY STATEMENT**
Data sharing is not applicable to this article as no new data were created or analyzed in this study.

**ORCID**
Wei Ma https://orcid.org/0000-0002-2952-7944
Hanzhong Liu https://orcid.org/0000-0002-0028-5136

**REFERENCES**
1. Yang L, Tsiatis AA. Efficiency study of estimators for a treatment effect in a pretest–Posttest trial. *Am Stat.* 2001;55(4):314-321.
2. Wang B, Ogburn EL, Rosenblum M. Analysis of covariance in randomized trials: more precision and valid confidence intervals, without model assumptions. *Biometrics.* 2019;75(4):1391-1400.
3. Freedman DA. On regression adjustments to experimental data. *Adv Appl Math.* 2008;40(2):180-193.
4. Lin W. Agnostic notes on regression adjustments to experimental data: reexamining Freedman’s critique. *Ann Appl Stat.* 2013;7:295-318.
5. Pocock SJ, Simon R. Sequential treatment assignment with balancing for prognostic factors in the controlled clinical trial. *Biometrics.* 1975;31(1):103-115.
6. Hu Y, Hu F. Asymptotic properties of covariate-adaptive randomization. *Ann Stat.* 2012;40(3):1794-1815.
7. Hu F, Ye X, Zhang LX. On the theory of multi-arm covariate-adaptive randomization. *Sci China Ser A.* 2022; in press.
8. Rosenberger WF, Lachin JM. *Randomization in Clinical Trials: Theory and Practice.* 2nd ed. New York, NY: John Wiley & Sons; 2015.
9. Lin Y, Zhu M, Su Z. The pursuit of balance: an overview of covariate-adaptive randomization techniques in clinical trials. *Contemp Clin Trials.* 2015;45:21-25.
10. Shao J, Yu X, Zhong B. A theory for testing hypotheses under covariate-adaptive randomization. *Biometrika.* 2010;97(2):347-360.
11. Shao J, Yu X. Validity of tests under covariate-adaptive biased coin randomization and generalized linear models. *Biometrics.* 2013;69(4):960-969.
12. Ma W, Hu F, Zhang LX. Testing hypotheses of covariate-adaptive randomized clinical trials. *J Am Stat Assoc.* 2015;110(510):669-680.
13. Zhu H, Hu F. Sequential monitoring of covariate-adaptive randomized clinical trials. *Stat Sin.* 2019;29:265-282.
14. Wang T, Ma W. The impact of misclassification on covariate-adaptive randomized clinical trials. *Biometrics.* 2021;77:451-464.
15. Bugni FA, Canay IA, Shaikh AM. Inference under covariate-adaptive randomization. *J Am Stat Assoc.* 2018;113(524):1784-1796.
16. Bugni FA, Canay IA, Shaikh AM. Inference under covariate-adaptive randomization with multiple treatments. *Quant Econ.* 2019;10:1747-1785.
17. Wang B, Susukida R, Mojtabai R, Amin-Esmaeili M, Rosenblum M. Model-robust inference for clinical trials that improve precision by stratified randomization and covariate adjustment. *J Am Stat Assoc.* 2021; in press.
18. Ye T, Shao J. Robust tests for treatment effect in survival analysis under covariate-adaptive randomization. *J R Stat Soc Ser B Stat Methodol.* 2020;82(5):1301-1323.
19. Liu H, Yang Y. Regression-adjusted average treatment effect estimators in stratified randomized experiments. *Biometrika.* 2020;107(4):935-948.
20. Lewis JA. Statistical principles for clinical trials. International conference on harmonisation E9 expert working group. Stat Med. 1999;15:1905-1942.
21. Committee for Medicinal Products for Human Use (CHMP). Guideline on Adjustment for Baseline Covariates in Clinical Trials. London: European Medicines Agency; 2015.
22. Neyman J, Dabrowska DM, Speed T. On the application of probability theory to agricultural experiments. Stat Sci. 1990;5:465-472.
23. Rubin DB. Estimating causal effects of treatments in randomized and nonrandomized studies. J Educ Psychol. 1974;66:688-701.
24. Rubin DB. Randomization analysis of experimental data: the Fisher randomization test comment. J Am Stat Assoc. 1980;75:591-593.
25. Wei LJ. An application of an urn model to the design of sequential controlled clinical trials. J Am Stat Assoc. 1978;73:559-563.
26. Zelen M. The randomization and stratification of patients to clinical trials. J Chronic Dis. 1974;27(7):365-375.
27. Efron B. Forcing a sequential experiment to be balanced. Biometrika. 1971;58(3):403-417.
28. Baldi Antognini A, Zagoraiou M. The covariate-adaptive biased coin design for balancing clinical trials in the presence of prognostic factors. Biometrika. 2011;98(3):519-535.
29. Cochran WG. Analysis of covariance: its nature and uses. Biometrics. 1957;13:261-281.
30. Cox DR, McCullagh P. Some aspects of analysis of covariance (with discussion). Biometrics. 1982;38:541-561.
31. Li X, Ding P. Rerandomization and regression adjustment. J R Stat Soc Ser B Stat Methodol. 2020;82:241-268.
32. Tsiatis AA, Davidian M, Zhang M, Lu X. Covariate adjustment for two-sample treatment comparisons in randomized clinical trials: a principled yet flexible approach. Stat Med. 2008;27(23):4658-4677.
33. Ye T, Yi Y, Shao J. Inference on the average treatment effect under minimization and other covariate-adaptive randomization methods. Biometrika. 2022;109(1):33-47.
34. Liu H, Tu F, Ma W. Lasso-adjusted treatment effect estimation under covariate-adaptive randomization. Biometrika. 2022; in press.
35. Keller MB, McCullough JP, Klein DN, et al. A comparison of nefazodone, the cognitive behavioral-analysissystem of psychotherapy, and their combination for the treatment of chronic depression. N Engl J Med. 2000;342(20):1462-1470.
36. Hilgers RD, Uschner D, Rosenberger WF, Heussen N. ERDO – A framework to select an appropriate randomization procedure for clinical trials. BMC Med Res Methodol. 2017;17:159.
37. Kuznetsova OM, Tymofyeyev Y. Preserving the allocation ratio at every allocation with biased coin randomization and minimization in studies with unequal allocation. Stat Med. 2012;31(8):701-723.
38. Liu Z, Yin J, Hu F. Covariate-adaptive designs with missing covariates in clinical trials. Sci China Ser A. 2015;58:1191-1202.
39. Liu Y, Hu F. Balancing unobserved covariates with covariate-adaptive randomized experiments. J Am Stat Assoc. 2022;117(538):875-886.
40. Li X, Zhou J, Hu F. Testing hypotheses under adaptive randomization with continuous covariates in clinical trials. Stat Methods Med Res. 2019;28:1609-1621.
41. Ma W, Qin Y, Li Y, Hu F. Statistical inference for covariate-adaptive randomization procedures. J Am Stat Assoc. 2020;115(531):1488-1497.
42. Zhang LX, Hu F, Cheung SH, Chan WS. Asymptotic properties of covariate-adjusted adaptive designs. Ann Stat. 2007;35:1166-1182.

SUPPORTING INFORMATION
Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Ma W, Tu F, Liu H. Regression analysis for covariate-adaptive randomization: A robust and efficient inference perspective. Statistics in Medicine. 2022;41(29):5645-5661. doi: 10.1002/sim.9585