Statistica Sinica Preprint No: SS-2025-0290	
Title	Efficient Covariate-adaptive Randomization in Clinical Trials with Smallest Selection Bias
Manuscript ID	SS-2025-0290
URL	http://www.stat.sinica.edu.tw/statistica/
DOI	10.5705/ss.202025.0290
Complete List of Authors	Li-Xin Zhang
Corresponding Authors	Li-Xin Zhang
E-mails	stazlx@mail.zjgsu.edu.cn
Notice: Accepted author version.	

Efficient Covariate-Adaptive Randomization with Smallest Selection Bias in Clinical Trials

Li-Xin Zhang

Zhejiang Gongshang University

Abstract: The power of conventional hypothesis tests and the selection bias in covariate-adaptive randomized clinical trials are typically investigated through simulation studies. In this article, we develop a theoretical framework for analyzing both the asymptotic power of linear-model-based tests for treatment effects and the asymptotic selection bias. Our results reveal that under covariate-adaptive randomization: (i) hypothesis tests generally lose power when covariates are not sufficiently balanced; in particular, the more covariates included in the testing model that are not accounted for in the randomization procedure, the greater the loss of power; (ii) the hypothesis test is usually more powerful than that under complete randomization; and (iii) compared with complete randomization, many popular covariate-adaptive randomization procedures in the literature—such as Pocock and Simon's marginal procedure, the stratified permuted block design, and Taves's minimization method—are generally efficient in terms of power but yield non-negligible selection bias. To address this trade-off, we propose a new family of covariate-adaptive randomization procedures that simultaneously account for both power and selection bias. Under these proce-

dures, covariate imbalances are kept sufficiently small to achieve asymptotically maximal power for testing treatment effects, while the selection bias remains asymptotically negligible. These theoretical results provide a comprehensive picture of how the power of hypothesis testing, covariate imbalance, and selection bias interact with one another.

Key words and phrases: Balancing covariates, clinical trial, loss of power, selection bias, Pocock and Simon's procedure

1. Introduction

It is well known that covariates often play an important role in clinical trials. Clinical trialists frequently express concern over unbalanced treatment arms with respect to key covariates of interest. To balance important covariates, covariate-adaptive designs (Rosenberger and Lachin, 2002) are usually employed. The most commonly implemented methods are stratified randomization and marginal minimization (McEntegart, 2003). Stratified randomization first stratifies the population and then applies separate randomization within each stratum; examples include the stratified permuted block design, among others. To handle many covariates, Taves (1974) and Pocock and Simon (1975) introduced the marginal minimization method, which attempts to minimize the weighted sum of differences of allocation numbers between treatment groups across all marginal values of covariates.

Hu and Hu (2012) discussed some limitations of these classical designs, proposed a generalized family of covariate-adaptive designs, and established their theoretical properties. For further discussion on handling covariates in clinical trials, see McEntegart (2003), Rosenberger and Sverdlov (2008), Hu and Rosenberger (2006), and references therein.

It is important to study the statistical inference associated with covariate-adaptive designs. In practice, conventional tests are often employed without accounting for the covariate-adaptive randomization scheme. It remains a concern whether conventional tests are still valid under covariate-adaptive designs, especially when the covariates used in trial design and those incorporated in inference procedures are not the same. A substantial body of simulation literature has examined statistical inference under covariate-adaptive designs. Birkett (1985) and Forsythe (1987) raised concerns about the validity of unadjusted analysis under covariate-adaptive designs and suggested, based on simulation studies, that all covariates used in Taves's minimization method should be included in the analysis to achieve a valid test. Feinstein and Landis (1976) and Green and Byar (1978) studied inference problems for stratified randomization with binary responses. Ciolino et al. (2011) showed that power loss could be non-negligible if the balancing of important continuous covariates is ignored even when adjustment is

made in the analysis for those covariates. More discussions can be found in Simon (1979), Tu, Shalay and Pater (2000), Aickin (2009) and elsewhere. Shao, Yu and Zhong (2010) and Ma, Hu and Zhang (2015) theoretically proved that, when some or all covariates used in trial design are not incorporated in inference procedures, the hypothesis test for comparing treatment effects is usually conservative in terms of Type I error. It is now generally accepted that covariates used in trial design should also be incorporated in inference procedures.

In this paper, we consider the opposite direction, namely, that covariates incorporated in inference procedures should be used in trial design. We aim to study the exact power of the hypothesis test for comparing treatment effects. An expansion of the relative power function is established. It is shown that when some or all covariates used in inference procedures are not incorporated in trial design, the test generally loses power; the more covariates in the testing model that are not incorporated in the randomization procedure, the greater the power loss. However, adaptive randomization is usually more powerful than complete randomization. When all the covariates incorporated in inference procedures are sufficiently balanced, the power of the test is asymptotically equivalent to the maximum achievable power. A broad class of covariate-adaptive designs, which includes most of

the covariate-adaptive designs in the literature, for example, Pocock and Simon's marginal procedure, the stratified permuted block design, can achieve the maximum power.

In the literature, continuous covariates are typically discretized in order to be included in the randomization scheme (Taves, 2010). However, as discussed in Scott et al (2002), breaking a continuous covariate into subcategories entails increased effort and loss of information. Ciolino et al. (2011) also pointed out that a lack of practical methods for balancing continuous covariates, combined with limited knowledge of the cost of failing to balance them, has led to a common phenomenon whereby continuous covariates are excluded from the randomization plan in clinical trials. Our theoretical results on power illustrate how discretization affects power loss and provide a formula for the cost of failing to balance continuous covariates.

On the other hand, randomization in clinical trials is fundamental to study design. Randomization is desirable for a number of reasons including the mitigation of selection bias which may arise if the person in charge of selecting patients has advance knowledge of the treatment assignments. Efron (1971) discussed how to measure the lack of randomness and treatment imbalance, and proposed a biased coin design (BCD) to trade off between the treatment imbalance and lack of randomness. However, the

lack of randomness of covariate-adaptive randomization has been seldom considered theoretically. Based on the measure of selection bias defined by Efron (1971), we find that, compared to complete randomization, the selection bias of both Pocock and Simon's marginal procedure and stratified permuted block design is non-negligible, even though the asymptotic power of these designs can be maximal. A natural question is whether there exists a covariate-adaptive randomization procedure under which both power and selection bias are simultaneously optimized. We propose a new family of covariate-adaptive designs that include Pocock and Simon's marginal procedure, Hu and Hu (2012)'s design, and Wei (1978)'s adaptive biased coin design. Within this family, a new covariate-adaptive randomization procedure is defined by choosing a suitable allocation function, under which covariate imbalances are small enough that the power of testing treatment effects is asymptotically maximal, while at the same time the asymptotic selection bias attains the minimum value of 1/2, so that every guessing strategy is asymptotically equally useless against the allocation procedure.

In Section 2, we study the power of the hypothesis tests for comparing the treatment effects. Section 3 addresses the theory of selection bias. In Section 4, the new family of adaptive randomization procedures are proposed. The proofs are provided in the Supplementary Material.

2. The power of hypothesis test under covariate-adaptive designs

In this section, we study the power of hypothesis testing based on a linear model framework for covariate-adaptive designs. Suppose two treatments 1 and 2 are studied under a covariate-adaptive randomized clinical trial, where μ_1 and μ_2 are parameters measuring the mean effects of treatment 1 and 2, respectively. Let n be the total number of patients enrolled in the study. Let T_i be the assignment of the ith patient, i.e., $T_i = 1$ for treatment 1 and $T_i = 0$ for treatment 2, $i = 1, 2, ..., n$. Conditional on the treatment assignment T_i, the following linear model is assumed for the response of the ith patient Y_i,

$$Y_i = \mu_1 T_i + \mu_2(1 - T_i) + \beta_1 X_{i,1} + ... + \beta_I X_{i,I} + \varepsilon_i, \tag{2.1}$$

where

- $X_{i,k}$s are discrete or continuous covariates which are independent and identically distributed as X_k , $k = 1, \ldots, I$;
- some or all values of $X_{i,k}$s are used in the randomization procedure, $k = 1, ..., I$;
- all covariates are independent of each other, and $EX_k = 0$ and $\mathrm{Var}(X_k) > 0$ for all k, $k = 1, \ldots, I$;

- ε_is are independent and identically distributed normal random errors with zero mean and variance σ_ε^2 and are independent of X_k,$k = 1, \ldots, I$.

Note that $X_{i,k}$ are assumed to be scalars in model (2.1). If $X_{i,k}$ is a discrete covariate, $X_{i,k}$ is a scalar that can take several values corresponding to different categories. In practice, a vector is usually used to represent a discrete covariate with multiple categories. In this paper, $X_{i,k}$ is assumed to be a scalar for simplicity, but all the results can be extended to the situation where discrete covariates with multiple categories are represented by vectors.

We let $\boldsymbol{Y} = (Y_1, Y_2, ..., Y_n)^T$, $\boldsymbol{\varepsilon} = (\varepsilon_1, \varepsilon_2, ..., \varepsilon_n)^T$, $\boldsymbol{\beta} = (\beta_1, ..., \beta_I)^T$, $\boldsymbol{\gamma} = (\mu_1, \mu_2, \beta_1, ..., \beta_I)^T$, and

$$\boldsymbol{X} = \begin{bmatrix} T_1 & 1-T_1 & X_{1,1} & \cdots & X_{1,I} \\ T_2 & 1-T_2 & X_{2,1} & \cdots & X_{2,I} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ T_n & 1-T_n & X_{n,1} & \cdots & X_{n,I} \end{bmatrix}.$$

Then model (2.1) can be written as $\boldsymbol{Y} = \boldsymbol{X}\boldsymbol{\gamma} + \boldsymbol{\varepsilon}$. The ordinary least squares method is used to obtain the estimate of $\boldsymbol{\gamma}$, which has the explicit form

$$\hat{\boldsymbol{\gamma}} = (\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{Y} = \boldsymbol{\gamma} + (\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{\varepsilon}.$$

In the covariate-adaptive randomization, the results of allocations $T_1, \ldots, T_n$ depend only on the covariates $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n$, and so are independent of $\boldsymbol{\epsilon}$. Hence, given $\boldsymbol{X}$, $\hat{\boldsymbol{\gamma}}$ follows a multi-dimensional normal distribution with mean $\boldsymbol{\gamma}$ and variance-covariance matrix $\sigma_\epsilon^2(\boldsymbol{X}^T\boldsymbol{X})^{-1}$.

When model (2.1) is constructed to study data from a covariate-adaptive randomized clinical trial, the primary interest is usually to compare treatment effects between different groups. To compare treatment effects of μ_1 and μ_2, the following right-sided hypothesis test is usually used

$$H_0 : \mu = 0 \text{ versus } H_A : \mu > 0. \tag{2.2}$$

where $\mu = \mu_1 - \mu_2$ is the treatment effect difference. The right-sided t-test statistic for (2.2) is

$$T = \frac{\boldsymbol{L}\hat{\boldsymbol{\gamma}}}{(\widehat{\sigma}_\epsilon^2 \boldsymbol{L}(\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{L}^T)^{1/2}}, \tag{2.3}$$

where $\boldsymbol{L} = (1, -1, 0, ..., 0)$, and $\hat{\sigma}_\epsilon^2 = (\boldsymbol{Y} - \boldsymbol{X}\hat{\boldsymbol{\gamma}})^T(\boldsymbol{Y} - \boldsymbol{X}\hat{\boldsymbol{\gamma}})/(n - I - 2)$ is the estimate of σ_ϵ^2. If $T > t_{1-\alpha}(\nu)$, where $t_{1-\alpha}(\nu)$ is the $(1-\alpha)$th quantile of a t-distribution with the degree of freedom $\nu = n - I - 2$, we reject the null hypothesis, otherwise we accept the null hypothesis. Given $\boldsymbol{X}$, the conditional power function is

$$\beta_{T,n}(\mu|\boldsymbol{X}) = 1 - F\left(t_{1-\alpha}(\nu); \nu, \frac{\mu}{\sigma_\epsilon^2 \boldsymbol{L}(\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{L}^T)^{1/2}}\right),$$

where $F(t;\nu,\delta)$ is the distribution function of a non-central t-distribution with non-central parameter δ and degree of freedom $\nu = n - I - 2$.

Let $\boldsymbol{X}_i = (X_{i,1}, \cdots, X_{i,I})^T$,

$$N_{n,1} = \sum_{i=1}^{n} T_i,\ N_{n,2} = \sum_{i=1}^{n}(1-T_i),\ D_n = N_{n,1} - N_{n,2},$$
$$\boldsymbol{D}_n^x = \sum_{i=1}^{n}(2T_i - 1)\boldsymbol{X}_i,\ D_n^{x_k} = \sum_{i=1}^{n}(2T_i - 1)X_{i,k},\ k = 1, \ldots, I,$$
$$\overline{\boldsymbol{X}}_{n,1} = \frac{\sum_{i=1}^{n} T_i \boldsymbol{X}_i}{N_{n,1}},\ \overline{\boldsymbol{X}}_{n,2} = \frac{\sum_{i=1}^{n}(1-T_i)\boldsymbol{X}_i}{N_{n,2}},$$
$$\overline{\boldsymbol{Y}}_{n,1} = \frac{\sum_{i=1}^{n} T_i \boldsymbol{Y}_i}{N_{n,1}},\ \overline{\boldsymbol{Y}}_{n,2} = \frac{\sum_{i=1}^{n}(1-T_i)\boldsymbol{Y}_i}{N_{n,2}},$$

$$\boldsymbol{S}_{n,xx} = \sum_{i=1}^{n} T_i(\boldsymbol{X}_i - \overline{\boldsymbol{X}}_{n,1})(\boldsymbol{X}_i - \overline{\boldsymbol{X}}_{n,1})^T + \sum_{i=1}^{n}(1-T_i)(\boldsymbol{X}_i - \overline{\boldsymbol{X}}_{n,2})(\boldsymbol{X}_i - \overline{\boldsymbol{X}}_{n,2})^T.$$

Here D_n is the difference between the numbers of patients in treatment group 1 and 2 overall, which can be regarded as a measure of overall treatment imbalance, and, $\boldsymbol{D}_n^x = \sum_{i=1}^{n} T_i \boldsymbol{X}_i - \sum_{i=1}^{n}(1-T_i)\boldsymbol{X}_i$ can be regarded as a measure of covariate imbalance. It can be seen that

$$\overline{\boldsymbol{X}}_{n,1} = \frac{n\overline{\boldsymbol{X}}_n + \boldsymbol{D}_n^x}{n + D_n} \text{ and } \overline{\boldsymbol{X}}_{n,2} = \frac{n\overline{\boldsymbol{X}}_n - \boldsymbol{D}_n^x}{n - D_n}$$

are functions of D_n and $\boldsymbol{D}_n^x$, where $\overline{\boldsymbol{X}}_n = \sum_{i=1}^n \boldsymbol{X}_i/n$. Also

$$\begin{aligned}\boldsymbol{L}(\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{L}^T =& \frac{1}{N_{n,1}} + \frac{1}{N_{n,2}} + (\overline{\boldsymbol{X}}_{n,1} - \overline{\boldsymbol{X}}_{n,2})^T \boldsymbol{S}_{xx}^{-1}(\overline{\boldsymbol{X}}_{n,1} - \overline{\boldsymbol{X}}_{n,2}) \\ =& \frac{4}{n\big(1-(D_n/n)^2\big)} + \frac{4Q_n^2}{n\big(1-(D_n/n)^2\big)^2},\end{aligned}$$

where

$$Q_n^2 = \big(\boldsymbol{D}_n^x/\sqrt{n} - \overline{\boldsymbol{X}}_n D_n/\sqrt{n}\big)^T \left(\frac{\boldsymbol{S}_{n,xx}}{n}\right)^{-1} \big(\boldsymbol{D}_n^x/\sqrt{n} - \overline{\boldsymbol{X}}_n D_n/\sqrt{n}\big).$$

Hence, the power function can be written as

$$\beta_{T,n}(\mu|\boldsymbol{X}) = 1 - F\left(t_{1-\alpha}(\nu); \nu, \frac{\mu}{2\sigma_\epsilon}\ell_n\right),$$

where

$$\ell_n = \sqrt{\frac{n\big(1-(D_n/n)^2\big)^2}{1-(D_n/n)^2 + \frac{1}{n}Q_n}} \le \sqrt{n}.$$

Because $F(t;\nu,\delta)$ is a non-increasing function of δ, it is obvious that when $D_n = 0$ and $\boldsymbol{D}_n^x = \boldsymbol{0}$, the power function takes its largest value

$$\beta_{T,n}(\mu|0) = 1 - F\left(t_{1-\alpha}(\nu); \nu, \frac{\mu}{2\sigma_\epsilon}\sqrt{n}\right).$$

A clinical trial is only a single realization of a random phenomenon, and it cannot be assumed that the observed imbalances D_n and $\boldsymbol{D}_n^x$ will be zero. Therefore, the loss of power is unavoidable. The problem is, compared to the largest power, how small is the realized power acceptable? The distance between the power functions $\beta_{T,n}(\mu|\boldsymbol{X})$ and $\beta_{T,n}(\mu|0)$ may reflect the loss

of power. However, since the power functions $\beta_{T,n}(\mu|\boldsymbol{X})$ and $\beta_{T,n}(\mu|0)$ are both convergent to 1, from their difference one can not discern how the imbalances D_n and $\boldsymbol{D}_n^x$ affect the loss of power clearly. To deal with this problem, we consider the relative loss of power as the ratio:

$$LossP_{T,n}(\mu|\boldsymbol{X}) = \frac{1-\beta_{T,n}(\mu|0)}{1-\beta_{T,n}(\mu|\boldsymbol{X})} = \frac{F\left(t_{1-\alpha}(\nu); \nu, \frac{u}{2\sigma_\epsilon}\sqrt{n}\right)}{F\left(t_{1-\alpha}(\nu); \nu, \frac{\mu}{2\sigma_\epsilon}\ell_n\right)},$$

$\nu = n - I - 2$. The smaller value of $LossP_{T,n}(\mu|\boldsymbol{X})$, the greater the power loss. The following theorem gives the order of the loss of power.

Theorem 1. *Write* $\sigma_{x,k}^2 = \mathrm{Var}(X_k)$, $k = 1, \ldots, I$,

$$V_n = (D_n/\sqrt{n})^2 + \sum_{k=1}^{I} \frac{(D_n^{x_k}/\sqrt{n})^2}{\sigma_{x,k}^2}.$$

Suppose $\frac{D_n}{n}$ *and* $\frac{\boldsymbol{D}_n^x}{n}$ *converge to zero in probability (or almost surely), then*

$$LossP_{T,n}(\mu|\boldsymbol{X}) = \exp\left\{-\frac{\mu^2}{8\sigma_\epsilon^2} V_n\big(1+o(1)\big)\right\}. \tag{2.4}$$

in probability (or almost surely).

In Theorem 1, the value of $V_n = (D_n/\sqrt{n})^2 + \sum_{i=1}^{p}(D_n^{x_k}/\sqrt{n})^2\sigma_{x,k}^{-2}$ gives the order of the loss of power. For complete randomization in which each patient is allocated to treatment 1 or 2 with the same probability 1/2,

$$\left(\frac{D_n}{\sqrt{n}}, \frac{D_n^{x_1}}{\sqrt{n}}, \cdots, \frac{D_n^{x_I}}{\sqrt{n}}\right)^T \xrightarrow{d} N_{I+1}\left(\mathbf{0}, diag(1, \sigma_{x,1}^2, \cdots, \sigma_{x,I}^2)\right),$$

where N_{I+1} is an $(I+1)$-dimensional normal distribution with mean and variance-covariance matrix as specified. So, $V_n \xrightarrow{d} \chi^2(I+1)$, where $\chi^2(I+1)$ is a χ^2-distributed random variable with $I+1$ degree of freedom. Hence we have

Corollary 1. *If the patients are allocated to treatments by the complete randomization, then*

$$LossP_{T,n}(\mu|\boldsymbol{X}) \xrightarrow{d} \exp\left\{-\frac{\mu^2}{8\sigma^2}\chi^2(I+1)\right\} < 1. \tag{2.5}$$

By (2.5), the expected relative loss of power is

$$\mathsf{E}\left[LossP_{T,n}(\mu|\boldsymbol{X})\right] \cong \left(1+\frac{\mu^2}{4\sigma_\epsilon^2}\right)^{-\frac{I+1}{2}}.$$

The right-side of (2.4) will converge to 1 whenever $V_n \to 0$. Therefore, we conclude the following corollary.

Corollary 2. *If*

$$\frac{D_n}{\sqrt{n}} \to 0 \quad \textit{and} \quad \frac{\boldsymbol{D}_n^x}{\sqrt{n}} \to 0 \tag{2.6}$$

in probability (or almost surely), then $LossP_{T,n}(\mu|\boldsymbol{X}) \to 1$ *in probability (or almost surely).*

Definition 1. When the condition (2.6) is satisfied, the power of the hypothesis test for comparing treatment effects under this allocation procedure is equivalent to that under the completely balanced allocation which

has the largest power. Such a randomization procedure is called *an efficient covariate-adaptive design*.

Condition (2.6) means that the covariates used in the testing model should be sufficiently balanced. In clinical trials, covariate-adaptive designs are usually based on discrete covariates (Taves 2010). If a continuous (or general) covariate is to be used in randomization, a discrete conversion need to be performed to break down the continuous covariate into a discrete variable with several subcategories. In general, the covariate-adaptive design is applied with respect to discrete variables $d_k(X_k)$, where d_ks are discrete functions. In such cases, define $\delta_k = X_k - E[X_k|d_k(X_k)]$ and $\tilde{X}_k = d_k(X_k)$. Accordingly, $\delta_{i,k} = X_{i,k} - E[X_{i,k}|d_k(X_{i,k})]$ and $\tilde{X}_{i,k} = d_k(X_{i,k})$ are ith observations of covariates δ_k and $\tilde{X}_k$, $k = 1, \ldots, I$. Suppose $\tilde{X}_k$ have m_k levels, resulting in $\prod_{k=1}^{I} m_k$ strata in total. Let $\tilde{\boldsymbol{X}}_i = (\tilde{X}_{i,1}, \ldots, \tilde{X}_{i,I})$ represent the covariate profile of the ith patient used in randomization, i.e., $\tilde{\boldsymbol{X}}_i = (\tilde{x}_1^{t_1}, \tilde{x}_2^{t_2}, \ldots, \tilde{x}_I^{t_I})$ if $\tilde{X}_{i,k}$ is at level $\tilde{x}_k^{t_k}$. For convenience, we use $(t_1, \ldots, t_I)$ to denote the stratum formed by patients who have the same covariate profile $(\tilde{x}_1^{t_1}, \ldots, \tilde{x}_I^{t_I})$, and use $(j; t_j)$ to denote the margin formed by patients with $\tilde{X}_j = \tilde{x}_j^{t_j}$. Then let

- $D_n = N_{n,1} - N_{n,2} = \sum_{i=1}^{n}(2T_i - 1)$ be the difference between the numbers of patients in treatment group 1 and 2 as total;

- $D_n(j;t_j) = \sum_{i=1}^{n}(2T_i - 1)I\{\tilde{X}_{i,j} = \tilde{x}_j^{t_j}\}$ be the differences between the numbers of patients in the two treatment groups on the margin $(j;t_j)$;

- $D_n(t_1,\ldots,t_I) = \sum_{i=1}^{n}(2T_i - 1)I\{\tilde{\boldsymbol{X}}_i = (\tilde{x}_1^{t_1},\ldots,\tilde{x}_I^{t_I})\}$ be the differences between the numbers of patients in the two treatment groups within the stratum $(t_1,\ldots,t_I)$.

These differences play important roles in the covariate-adaptive randomization procedure and in properties of statistical inference for covariate-adaptive designs. Each one of these differences is used to measure the *imbalance* at the corresponding level (overall, marginal, or within-stratum) of the n assignments.

Theorem 2. *Let* $\sigma^2_{\delta,k} = E[Var(\delta_k|d_k(X_k))] = E[\delta_k^2]$, $k = 1,\ldots,p$. *Assume that each assignment* T_m *depends only on* $T_1,\ldots,T_{m-1}$ *and* $\tilde{\boldsymbol{X}}_1,\ldots,\tilde{\boldsymbol{X}}_m$, *i.e., the probability of* $T_m = 1$ *is a function of* $T_1,\ldots,T_{m-1}$ *and* $\tilde{\boldsymbol{X}}_1,\ldots,\tilde{\boldsymbol{X}}_m$. *Suppose*

$$\frac{D_n}{\sqrt{n}} \to 0 \quad \textit{and} \quad \frac{D_n(j;t_j)}{\sqrt{n}} \to 0 \tag{2.7}$$

in probability (or almost surely), $t_j = 1,\ldots,m_j$, $j = 1,\ldots,I$. *Then*

$$LossP_{T,n}(\mu|\boldsymbol{X}) \xrightarrow{d} \exp\left\{-\frac{\mu^2}{8\sigma_\epsilon^2}\sum_{k=1}^{I}\frac{\sigma^2_{\delta,k}}{\sigma^2_{x,k}}\chi^2_k(1)\right\}, \tag{2.8}$$

where $\chi^2_1(1),\ldots,\chi^2_I(1)$ *are independent* χ^2*-distributed random variables with*

1 *degree of freedom. The expected relative loss of power is*

$$E[LossP_{T,n}(\mu|\boldsymbol{X})] \cong \prod_{i=1}^{I}\left(1+\frac{\mu^2}{4\sigma_\epsilon^2}\frac{\sigma_{\delta,k}^2}{\sigma_{x,k}^2}\right)^{-1/2}.$$

Condition (2.7) is satisfied under various covariate-adaptive designs for balancing covariates. For example, under stratified permuted block designs, the class of covariate-adaptive designs proposed by Hu and Hu (2012), and Pocock and Simon (1975)'s marginal procedure,

$$D_n \text{ and } D_n(j;t_j), \text{ are bounded in probability,} \tag{2.9}$$

$t_j = 1,\ldots,m_j$, $j = 1,\ldots,I$ (c.f. Ma, Hu and Zhang (2015); Hu, Ye and Zhang (2023)), and so (2.7) is satisfied. Further, if the covariates X_ks are discrete and $d_k(X_k) = X_k$, $k = 1,\ldots,I$, then D_n and $\boldsymbol{D}_n^x$ are bounded in probability, resulting in $V_n = O(n^{-1})$ and

$$V_n = O(n^{-1}) \text{ and } LossP_{T,n}(\mu|\boldsymbol{Z}) = 1 + O(n^{-1}) \text{ in probability.}$$

Notice that $O(n^{-1})$ is the fastest convergence rate unless the treatments are completely balanced so that $D_n = 0$ and $D_n(j,t_j) = 0$ for all t_j and j, which indicates that Efron's biased coin design, as well as its generalizations such as Pocock and Simon's marginal procedure and Hu and Hu's design, is usually efficient in terms of power with respect to the problem it addresses.

Simulation studies have shown that Taves (1974)'s minimization method also reduces marginal imbalances as well as the overall imbalance. However,

theoretical evidence has not, to the best of our knowledge, been found in literature. Whether (2.9) holds or not is still an open problem. We will show that (2.7) is true under Taves's minimization method (c.f., Theorem 4 and Corollary 3), so it is an efficient covariate-adaptive design in the sense that $LossP_{T,n}(\mu|\boldsymbol{X}) \xrightarrow{P} 1$ if the covariates X_ks are discrete and $d_k(X_k) = X_k$, $k = 1, \ldots, I$.

In Theorem 2, the ratio $\sigma^2_{\delta,k}/\sigma^2_{x,k}$ describes the cost to loss of power of failing to balance the kth covariate completely. It is obvious that $0 \leq \sigma^2_{\delta,k}/\sigma^2_{x,k} \leq 1$. The larger $\sigma^2_{\delta,k}$ is, the larger the cost is. If $\mathsf{E}[X_k|d_k(X_k)] = X_k$, i.e., the k-covariate is not considered in the randomization, then $\sigma^2_{\delta,k}/\sigma^2_{x,k} = 1$ and accordingly, the k-covariate contributes fully to the loss of power and so the cost is the largest. If $\mathsf{E}[X_k|d_k(X_k)] = d_k(X_k)$, i.e., the kth covariate is discrete and is balanced enough with respect to all of its values, then $\sigma^2_{\delta,k} = 0$ and accordingly, the contribution of the k-covariate to the loss of power can be neglected. So, the fewer covariates and their values that are balanced, the greater the power loss. Comparing the results in Corollary 1 and Theorem 2, we find that the hypothesis test for comparing treatment effects under a covariate-adaptive randomization procedure is usually more powerful than the one under the complete randomization.

In the literature, covariate-adaptive randomization schemes are usually

based on discrete covariates. When the covariates are continuous, the covariates are usually discretized. However, as discussed in Scott et al (2002), the discretization means increased effort and loss of information. Ciolino et al. (2011) also pointed out that uncontrolled imbalances in continuous covariates may affect the statistical analysis of the trial outcome. In the formula (2.8), $\sigma^2_{\delta,k}$ is the variance of the uncontrolled part of the kth covariate. The ratio $\sigma^2_{\delta,k}/\sigma^2_{x,k}$ just gives the cost of failing to balance the covariate X_k completely. It is easily seen that the cost can be decreased by refining the discretization. Recently, Ma et al. (2024) proposed a procedure to balance the continuous covariates $\boldsymbol{X}_i$s directly so that the condition (2.6) is satisfied under widely satisfied conditions. Under Ma et al. (2024)'s procedure, the asymptotic power is the largest when the covariates considered in the randomization and the covariates used in the hypothesis test for treatment effects are consistent.

The inconsistency of the covariates considered in the randomization and the covariates used in the hypothesis test for treatment effects may affect the hypothesis test. The evidences have been found in many simulation studies (Birkett, 1985; Forsythe, 1987; Tu, Shalay and Pater, 2000; Aickin, 2009; Ciolino et al. , 2011). Ma, Hu and Zhang (2015), Zhu et al (2025) etc derived the theory for the case where the covariates considered

in the testing are fewer than those considered in the randomization under a large class of covariate-adaptive randomization procedures satisfying (2.7). Combining the results in Theorem 2 and those of Ma, Hu and Zhang (2015), we conclude that, under a large class of covariate-adaptive designs for balancing covariates, (i) if the covariates considered in the testing are less than those considered in the randomization, then the hypothesis test is usually conservative in terms of Type I error, and, the more of the values of covariates are not consistent, the more the testing is conservative; (ii) if the covariates considered in the testing are more than those considered in the randomization, then the hypothesis test usually loses power, and, the more of the values of covariates are not consistent, the more the power loss; (iii) in any case, the hypothesis test is usually more powerful than the one under complete randomization.

Remark 1. The conclusions of Theorems 1 and 2 remain true when the statistic $|T|$ is used to test the following two-sided hypothesis test

$$H_0 : \mu = 0 \text{ versus } H_A : \mu \neq 0. \tag{2.10}$$

When σ_ϵ^2 is known, Theorems 1 and 2 are also valid for one-sided or two-sided z-test.

The formula (2.4) remains true under the alternative hypothesis with

$0 < \mu = \mu_n \to 0$ and $\mu_n\sqrt{n} \to \infty$. But it no longer holds when a sequence of local alternative hypotheses of the type $H_A : \mu = \eta/\sqrt{n} > 0$ are considered. In such case, one can show that $\frac{d}{d\delta}F(t;\nu,\delta)$ converges to a continuous function $f_\infty(t;\delta) < 0$ as $\nu \to \infty$ uniformly in t and δ on bounded intervals. It follows that

$$\begin{aligned} &F\Big(t_{1-\alpha}(\nu);\nu,\frac{\eta}{2\sigma_\epsilon}\frac{\ell_n}{\sqrt{n}}\Big) - F\Big(t_{1-\alpha}(\nu);\nu,\frac{\eta}{2\sigma_\epsilon}\Big) \\ =& -\frac{\eta}{4\sigma_\epsilon}f_\infty\Big(z_{1-\alpha};\frac{\eta}{2\sigma_\epsilon}\Big)\frac{n-\ell_n^2}{n}\big(1+o(1)\big) = C_{\alpha,\eta}\frac{V_n}{n}\big(1+o(1)\big). \end{aligned}$$

Hence

$$\beta_{T,n}(u|0) - \beta_{T,n}(u|\boldsymbol{X}) = O(1)\frac{V_n}{n}. \tag{2.11}$$

The value of V_n also reflects the order of power loss.

When the distribution of the model error ϵ_i is not normal, though it is difficult to derive the exact power function, (2.11) is expected to be true especially when the errors follow a continuous distribution. An approximation method is usually applied to approximate the true distribution of the statistic T for hypothesis testing, when the distribution of error ϵ_i is not normal or unknown. It is well known that if normal approximation is used, the approximation precision of the distribution as well as the power function is $O\left(1/\sqrt{n}\right)$. If the Bootstrap method is used, the approximation precision would be improved to $O\left(1/n\right)$. Hence, compared to the approximation pre-

cision, the loss of power $O(1)\frac{V_n}{n}$ can be ignored whenever $V_n \to 0$. Again, the condition (2.6) is needed.

3. Selection bias

As advocated in Efron (1971), selection bias (lack of randomness) in a design manifests in the possibility for the experimenter to partially guess the sequence of treatment allocations; thus it can be measured, say, by the expected percentage of correct guesses when an optimal guessing strategy is used. Let $J_m = 1$ if the mth assignment is guessed correctly, and $J_m = 0$ otherwise. The expected proportion of correct guesses is

$$SB_n(\Delta) = \mathsf{E}\left[\frac{1}{n}\sum_{m=1}^{n} J_m\right] = \frac{1}{n}\sum_{m=1}^{n} \mathsf{P}(J_m = 1).$$

Here $SB_n(\Delta)$ stands for the selection bias of the allocation procedure Δ. Every guessing strategy is equally useless against complete randomization, yielding the expected percentage of correct guesses $\frac{1}{2}$ for any n; this is the optimal value of SB_n.

For adaptive randomization, the optimal strategy consists, at each step, of picking the under-represented treatment, which is always the treatment with the higher probability of being allocated, with no preference in case of a tie. Let p_m be the conditional probability of assigning the mth patient to treatment 1 given the history σ-field including the information of past

allocations, past covariates and current covariate observed. The probability that the mth assignment is guessed correctly is $\mathsf{E}[p_m \vee (1-p_m)]$. So,

$$SB_n(\Delta) = \frac{1}{n}\sum_{m=1}^{n} \mathsf{E}\left[p_m \vee (1-p_m)\right] = \frac{1}{2} + \frac{1}{n}\sum_{m=1}^{n} \mathsf{E}\Big[\big|p_m - \frac{1}{2}\big|\Big].$$

There are other ways to measure the selection bias or predictability of a randomization procedure. For example, an equivalent way is to consider Simth (1984)'s index, namely the difference between the mean percentage of correct guesses and that of incorrect guesses: $U_n(\Delta) = 2SB_n(\Delta) - 1$, which has the advantage of ranging from 0 to 1. Also, the selection bias can be expressed as $\frac{1}{n}\sum_{m=1}^{n} \mathsf{E}[|p_m - 1/2|]$, which is just the Simth (1984)'s index $U_n(\Delta) = 2SB_n(\Delta) - 1$. For the discussion on the predictability and selection bias in randomized block procedure, one can refer to Berger, Ivanova, and Knoll (2003). In this paper, we only consider Efron's definition. It is easily seen that $SB_n(\Delta) > 1/2$ unless $p_m \equiv 1/2$ for all m. In general, we consider the asymptotic selection bias $SB(\Delta) = \lim_{n\to\infty} SB_n(\Delta)$.

Under stratified permuted block designs, or Taves (1974)'s minimization method, a positive percentage of patients are assigned deterministically, and so are guessed correctly, resulting in

$$\liminf_{n\to\infty} SB_n > 1/2.$$

For the class of biased coin designs, when the design does not include the

covariate,Efron (1971) gives the formula for the asymptotic selection bias SB of his original biased coin design, which indicates that $SB(BCD) > 1/2$.

For Pocock and Simon (1975)'s procedure, no closed form of the asymptotic selection bias SB has been found in the literature. But we have the following result.

Theorem 3. *For the Pocock and Simon's procedure, Hu and Hu (2012)'s procedure or more general procedures as in Remark 2, we have*

$$\lim_{n\to\infty} SB_n > \frac{1}{2}.$$

So, for all the covariate-adaptive randomization procedures that have been introduced so far, when the condition (2.7) is satisfied so that the loss of power would be negligible, the asymptotic selection bias SB is larger than $1/2$ so that the lack of randomness is non-negligible.

It is obvious that, if

$$p_m \to \frac{1}{2} \quad \text{in probability,} \tag{3.1}$$

then SB attains the minimum value $1/2$ of the selection biases, and every guessing strategy is asymptotically equally useless against the design. When the trials do not include the covariate, Wei (1978) defines the adaptive BCD satisfying (3.1) such that $SB = 1/2$. But Wei's design does not satisfy (2.7).

The problem is whether we can find a randomization procedure such that $SB = 1/2$ and the condition (2.7) is also satisfied? The next section will deal with this problem.

4. A New family of Covariate-Adaptive Randomization methods

In this section, we consider the randomization methods. We first give a general framework for covariate-adaptive randomization. We consider the same setting as that of Pocock and Simon (1975) and only focus on two treatment groups 1 and 2 here. As before, let T_j be the assignment of the jth patient, $j = 1, \ldots, n$, i.e., $T_j = 1$ for treatment 1 and $T_j = 0$ for treatment 2. Recall that $\tilde{\boldsymbol{X}}_j = \big(d_1(X_{j,1}), \ldots, d_I(X_{j,I})\big)$ indicates the discrete part of the covariate profile of that patient considered in adaptive randomization, i.e., $\tilde{\boldsymbol{X}}_j = (\tilde{x}_1^{t_1}, \ldots, \tilde{x}_I^{t_I})$ if his or her ith covariate is at level $\tilde{x}_i^{t_i}$, $1 \le i \le I$ and $1 \le t_i \le m_i$. For convenience, we use $(t_1, \ldots, t_I)$ to denote the *stratum* formed by patients who possess the same covariate profile $(\tilde{x}_1^{t_1}, ..., \tilde{x}_I^{t_I})$, resulting in $M = \prod_{i=1}^{I} m_i$ strata, and use $(i; t_i)$ to denote the *margin* formed by patients whose ith covariate is at level $\tilde{x}_i^{t_i}$. Recall that $D_m = \sum_{j=1}^{m}(2T_j - 1) = N_{m,1} - N_{m,2}$, $D_m(i; t_i) = \sum_{j=1}^{m}(2T_j - 1)I\{\tilde{X}_{j,i} = \tilde{x}_i^{t_i}\}$ and $D_m(t_1, \ldots, t_I) = \sum_{j=1}^{m}(2T_j - 1)I\{\tilde{\boldsymbol{X}}_j = (\tilde{x}_1^{t_1}, \ldots, \tilde{x}_I^{t_I})\}$ are measures of the overall, marginal and within-stratum imbalances of the

first m assignments, respectively.

4.1 The randomization procedure

The procedure is defined as follows:

1) The first patient is assigned to treatment 1 with probability 1/2.

2) Suppose $(n-1)$ patients have been assigned to treatments $(n > 1)$ and the covariate value $\boldsymbol{X}_n = \tilde{\boldsymbol{x}}_n^* = (\tilde{x}_1^{t_1^*}, \ldots, \tilde{x}_I^{t_I^*})$ of the nth patient is observed and falls within stratum $(t_1^*, \ldots, t_I^*)$.

   For the first $(n-1)$ patients, let D_{n-1}, $D_{n-1}(i; t_i^*)$ and $D_{n-1}(t_1^*, \ldots, t_I^*)$ be the overall imbalance, the marginal imbalance for the margin $(i; t_i^*)$ and the within-stratum imbalance for the stratum $(t_1^*, \ldots, t_I^*)$, respectively.

3) If the nth patient were assigned to treatment 1, then $D_n^{(1)} = D_{n-1} + 1$ would be the "potential" overall difference in the two groups; similarly,

$$D_n^{(1)}(i; t_i^*) = D_{n-1}(i; t_i^*) + 1$$

   and

$$D_n^{(1)}(t_1^*, \ldots, t_I^*) = D_{n-1}(t_1^*, \ldots, t_I^*) + 1$$

   would be the potential differences on margin $(i; t_i^*)$ and within stratum $(t_1^*, \ldots, t_I^*)$, respectively.

4) Define an imbalance measure $Imb_n^{(1)}$ by

$$Imb_n^{(1)} = w_o[D_n^{(1)}]^2 + \sum_{i=1}^{I} w_{m,i}[D_n^{(1)}(i;t_i^*)]^2 + w_s[D_n^{(1)}(t_1^*,\ldots,t_I^*)]^2,$$

which is the weighted imbalance that would be caused if the nth patient were assigned to treatment 1. w_o, $w_{m,i}$ $(i = 1,\ldots,I)$ and w_s are nonnegative weights placed on overall, within a covariate margin and within a stratum cell, respectively. Without loss of generality we can assume

$$w_o + w_s + \sum_{i=1}^{I} w_{m,i} = 1.$$

5) In the same manner we can define $Imb_n^{(2)}$, the weighted imbalance that would be caused if the nth patient were assigned to treatment 2. In this case, the three types of potential differences are the existing ones minus 1, instead of plus 1.

6) Conditional on the assignments of the first $(n-1)$ patients as well as the covariates' profiles of the first n patients, assign the nth patient to treatment 1 with probability

$$\begin{aligned} &P(T_n = 1|\boldsymbol{Z}_{n-1}, \tilde{\boldsymbol{X}}_n = (\tilde{x}_1^{t_1^*},\ldots,\tilde{x}_I^{t_I^*}), \boldsymbol{T}_{n-1}) \\ &= g_n\left(Imb_n^{(1)} - Imb_n^{(2)}\right) \end{aligned} \tag{4.1}$$

where $n > 1$, $\boldsymbol{Z}_{n-1} = (\tilde{\boldsymbol{X}}_1,\ldots,\tilde{\boldsymbol{X}}_{n-1})$, $\boldsymbol{T}_{n-1} = (T_1,\ldots,T_{n-1})$. Here

$g_n(x)$ is a real function with $0 \leq g_n(x) \leq 1$. It is called an allocation function.

Using the basic equation $(x+1)^2 - (x-1)^2 = 4x$, the critical quantity $Imb_n^{(1)} - Imb_n^{(2)}$ in Step 6) can be simplified as

$$\begin{aligned} & Imb_n^{(1)} - Imb_n^{(2)} \\ =& 4\left\{ w_o D_{n-1} + \sum_{i=1}^{I} w_{m,i} D_{n-1}(i; t_i^*) + w_s D_{n-1}(t_1^*, \ldots, t_I^*) \right\} \\ :=& 4 \cdot \Lambda_{n-1}(t_1^*, \ldots, t_I^*) \end{aligned} \tag{4.2}$$

Therefore, the allocation probability $g_n\left(Imb_n^{(1)} - Imb_n^{(2)}\right)$ is determined by the value of $\Lambda_{n-1}(t_1^*, \ldots, t_I^*)$, which is a weighted average of current imbalances at different levels.

This framework includes various kinds of covariate-adaptive randomization procedures. If stratified randomization is considered, one need only to let $w_s = 1$ and the other weights be zero. If $w_o = w_s = 0$, the design is a marginal randomization procedure.

4.2 The choice of the allocation function

In the literature different views have been given as to the selection of the allocation probability function $g_n(\cdot)$. Efron (1971), Pocock and Simon (1975), Hu and Hu (2012) suggested

$$g_n(x) = 1 - \rho \text{ if } x > 0, \ \frac{1}{2} \text{ if } x = 0, \ \rho \text{ if } x < 0, \tag{4.3}$$

where $1/2 < \rho < 1$. If the allocation function is chosen in this way, and $w_s = w_o = 0$, then the design is the Pocock and Simon (1975)'s marginal procedure. Ma, Hu and Zhang (2015) theoretically proved that under Pocock and Simon's procedure, the marginal imbalances and overall imbalance are bounded in probability. But the selection bias is not negligible as we have seen in Section 3.

When the design does not include the covariate, Wei (1978) defines the adaptive BCD with allocation function defined as

$$g_n(x) = g\left(\frac{x}{n-1}\right), \tag{4.4}$$

where $g(x)$ is a non-increasing function with $0 \leq g(x) \leq 1$, $g(x) = 1 - g(-x)$, $g(0) = 0.5$ and $g'(0) < 0$. It has been shown that $SB(WeiBCD) = 1/2$, which attains the minimum value as complete randomization does. However, Wei's BCD does not satisfy the condition (2.7) and so the power loss could be non-negligible (Antognini and Giovagnoli, 2004; Antognini, 2008).

Next, we show that a covariate-adaptive design can be defined by choosing suitable allocation function $g_n(x)$ so that the selection bias is asymptotically at its minimum and the covariate imbalances considered are of the order of $o(n^{1/2})$ in probability for which the loss of power would be negligible.

In general, we assume the allocation function $g_n(x)$ $(0 \le g_n(x) \le 1)$ satisfies the following conditions

$$g_n(x) \le 1/2 \le g_n(-x) \text{ when } x \ge 0, \tag{4.5}$$

$$\frac{|1/2 - g_n(x_n)|}{|x_n|/n} \to +\infty \text{ whenever } 0 < \frac{|x_n|}{n} \to 0 \tag{4.6}$$

and

$$g_n(x_n) \to \frac{1}{2} \text{ whenever } \frac{x_n}{n} \to 0. \tag{4.7}$$

We will show that if the allocation function satisfies the conditions (4.5) and (4.6), then covariate imbalances are of the order of $o(n^{1/2})$ in probability and, if the conditions (4.5) and (4.7) are satisfied, then the asymptotic selection bias is 1/2. The allocation function $g_n(x)$ of BCD given in (4.3) satisfies (4.5) and (4.6), but it does not satisfy condition (4.7). Wei's allocation function $g_n(x)$ given in (4.4) satisfies (4.5) and (4.7) , but it does not satisfy condition (4.6).

It is easily seen that $g_n(x) = 1 - \Phi\left(sgn(x)\sqrt{|x|/n}\right)$ satisfies all conditions (4.5)-(4.7), where $\Phi(x)$ is the standard normal distribution function. It can be chosen as an allocation function to define a design as desired.

In the following we will give a specific allocation function. For this allocation function, the design has many fine properties. Let $0 \le g(x) \le 1$

be a non-increasing function with $g(0) = 0.5$, $g'(0) < 0$. Define

$$g_n(x) = g\left(\frac{x}{(n-1)^\gamma}\right), \quad 0 \le \gamma \le 1. \tag{4.8}$$

When $\gamma = 1$, the allocation function $g_n(x)$ is Wei's function given in (4.4). When $\gamma = 0$, the design is a generalization of the Pocock and Simon (1975)'s procedure as well as Hu and Hu (2012)'s design. In this paper, we mainly consider the case $0 < \gamma < 1$.

4.3 Theoretical Properties

To consider the asymptotic properties of the design, we need some more notations. For the first n patients, we know that $D_n(t_1, \ldots, t_I)$ is the true difference between the two treatment arms within stratum $(t_1, \ldots, t_I)$. Let

$$\boldsymbol{D}_n = [D_n(t_1, \ldots, t_I)]_{1 \le t_1 \le m_1, \ldots, 1 \le t_I \le m_I}$$

be an array of dimension $m_1 \times \ldots \times m_I$ which stores the current assignment differences in all strata and so stores the current imbalances. Also, the covariates $\tilde{\boldsymbol{X}}_1, \tilde{\boldsymbol{X}}_2, \ldots$ are independently and identically distributed. Since $\tilde{\boldsymbol{X}}_n = (\tilde{x}_1^{t_1}, ..., \tilde{x}_I^{t_I})$ can take $M = \prod_{i=1}^{I} m_i$ different values, it in fact follows an M-dimension multinomial distribution with parameter $\boldsymbol{p} = (p(t_1, \ldots, t_I))$, where each element $p(\boldsymbol{t}) = \mathsf{P}(\tilde{\boldsymbol{X}}_n = (\tilde{x}_1^{t_1}, \ldots, \tilde{x}_I^{t_I}))$ is the probability that a patient falls within the corresponding stratum. Without loss of generality, we assume $p(t_1, \ldots, t_I) > 0$ for all $(t_1, \ldots, t_I)$.

Write $\boldsymbol{t} = (t_1, \ldots, t_I)$. Define

$$M_n = \sum_{\boldsymbol{t}} w_s D_n^2(\boldsymbol{t}) + \sum_{i=1}^{I} \sum_{t_i=1}^{m_i} w_{m,i} D_n^2(i; t_i) + w_o D_n^2.$$

Now we give our main results.

Theorem 4. *Suppose the allocation function $g_n(x)$ satisfies the conditions* (4.5) *and* (4.6). *Then*

$$E\left(\frac{M_n}{n}\right)^r \to 0 \ \ \forall r > 0, \tag{4.9}$$

i.e., $M_n = o(n)$ in L_r for all $r > 0$. In particular,

(i) *If $w_s > 0$, then $\boldsymbol{D}_n = o(n^{1/2})$ in L_r for any $r > 0$;*

(ii) *If $w_s + w_{m,i} > 0$, then $D_n(i; t_i) = o(n^{1/2})$ in L_r for any $r > 0$, $t_i = 1, \ldots, m_i$;*

(iii) *For any case $D_n = o(n^{1/2})$ in L_r for any $r > 0$.*

Further, if the condition (4.7) *is satisfied, then $SB_n \to \frac{1}{2}$ and $\boldsymbol{D}_n = o(n)$ in L_r for any $r > 0$.*

Theorem 5. *Let the allocation function $g_n(x)$ be defined as* (4.8) *with $0 < \gamma \leq 1$, and $0 \leq g(x) \leq 1$ be a non-increasing function with $g(0) = 1/2$, $g'(0) < 0$. Then*

$$M_n = O(n^{\gamma}) \quad \text{in } L_r \text{ for all } r > 0, \tag{4.10}$$

$$M_n = o(n^{\gamma+\epsilon}) \text{ a.s. for any } \epsilon > 0 \tag{4.11}$$

and

$$SB_n = \frac{1}{2} + O(n^{-\gamma/2}). \tag{4.12}$$

In particular,

(i) *If* $w_s > 0$, *then* $\boldsymbol{D}_n = O(n^{\gamma/2})$ *in* L_r *for any* $r > 0$, *and* $\boldsymbol{D}_n = o(n^{\gamma/2+\epsilon})$ *a.s. for any* $\epsilon > 0$;

(ii) *If* $w_s + w_{m,i} > 0$, *then* $D_n(i; t_i) = O(n^{\gamma/2})$ *in* L_r *for any* $r > 0$, *and* $D_n(i; t_i) = o(n^{\gamma/2+\epsilon})$ *a.s. for any* $\epsilon > 0$, $t_i = 1, \ldots, m_i$;

(iii) *For any case* $D_n = O(n^{\gamma/2})$ *in* L_r *for any* $r > 0$, *and* $D_n = o(n^{\gamma/2+\epsilon})$ *a.s. for any* $\epsilon > 0$, $\boldsymbol{D}_n = o(n)$ *a.s. and in* L_r *for any* $r > 0$;

(iv) *Suppose* $w_s + w_{m,i} > 0$, $X_i = d_i(X_i)$ *(the covariates in the test and in the randomization procedure are the same),* $i = 1, \ldots, I$. *Then*

$$Loss P_{T,n}(\mu|\boldsymbol{X}) = 1 + O\big(n^{\gamma-1}\big) \quad \text{in probability.}$$

Further, in (4.12), $O(n^{-\gamma/2})$ *can be written as* $a_n n^{-\gamma/2}$ *with*

$$c_0 < \liminf_{n\to\infty} a_n \le \limsup_{n\to\infty} a_n \le \frac{2\sqrt{|g'(0)|}}{2-\gamma},$$

where $c_0 > 0$ *is a constant which does not depend on* γ.

The value of γ gives the order of the allocation bias as well as the order of selection bias, which provides a picture of how the allocation bias and the selection bias conflict with each other. The smaller γ is, the smaller the allocation bias is, but the larger the selection bias is. In practice, one may choose $\gamma = 1/2$ to give a tradeoff between the two types of bias, both of which are moderate.

Remark 2. Theorem 5 does not include the case of $\gamma = 0$. When $\gamma = 0$, suppose $0 < g(x) \leq 1/2 \leq g(-x) < 1$ for all $x \geq 0$, and $\liminf_{x\to\infty} |g(x)-1/2| > 0$. It can be shown that $[\Lambda_n(t_1, \ldots, t_I)]_{t_i=1,\ldots,m_i, i=1,\ldots,I}$ is a positive recurrent Markov chain with an invariant distribution $\boldsymbol{\pi}$ and $\sup_n \mathsf{E}\,[M_n^r] < \infty$ for all $r > 0$ (c.f. Hu and Zhang (2020)). Further,

$$\lim_{n\to\infty} SB_n = \frac{1}{2} + \sum_{\boldsymbol{t}} \mathsf{E}_{\boldsymbol{\pi}}\Big[\Big|g\big(\Lambda(\boldsymbol{t})\big) - \frac{1}{2}\Big|\Big] p(\boldsymbol{t}) > 1/2.$$

So, Theorem 5 is still valid for $\gamma = 0$.

It should be noticed that Taves (1974)'s minimization method is not included in Remark 2, because the related allocation function is $g_n(x) = 1/2$ if $x = 0$, $g_n(x) = 0$ if $x > 0$ and $g_n(x) = 1$ if $x < 0$. This allocation function $g_n(x)$ satisfies the conditions (4.5) and (4.6). So (4.9) and (i)-(iii) of Theorem 4 remain true. We have more precise results as in the following corollary.

Corollary 3. *Suppose the allocation function $g_n(x)$ satisfies the conditions* (4.5) *and there exist constants $x_0 > 0$ and $p_0 > 0$ such that*

$$\left|g_n(x) - \frac{1}{2}\right| \geq p_0 \quad \textit{for all } |x| \geq x_0 \textit{ and } n \geq 1. \tag{4.13}$$

Then

$$M_n = O(n^{\epsilon}) \ a.s. \quad \textit{and} \quad \textit{in } L_r \textit{ for all } r > 0, \epsilon > 0. \tag{4.14}$$

5. Concluding Remarks

In this paper, we have studied the theoretical properties of hypothesis testing and selection bias under covariate-adaptive designs. First, based on a linear normal model framework, we derive the corresponding asymptotic power of the hypothesis test for comparing treatment effects. The asymptotic relative power is a decreasing function of the covariate imbalances, characterizing the impact of covariate imbalance on statistical power. Second, it is shown that, for many classical covariate-adaptive randomization procedures including stratified permuted block designs, Taves (1974)'s minimization method, Pocock and Simon (1975)'s procedure, and Hu and Hu (2012)'s procedure, the selection bias is non-negligible. Finally, a new family of covariate-adaptive designs are proposed under which the power of testing the treatment effects is asymptotically the largest and, at the same time, the asymptotic selection bias attains the minimum value.

To apply Theorem 2 on the asymptotic power to a specific covariate-adaptive randomized clinical trial and find an efficient randomization, one only needs to check the condition (2.7) or (2.6). The conclusion can be applied to a broad range of designs. To check or find the most random covariate-adaptive randomization procedure, one only needs to check a simple condition (3.1). The new family of covariate-adaptive randomization procedures are shown to satisfy the condition (2.7) and (3.1) simultaneously. The results in this paper provide new insights about balance, selection bias, efficiency, and randomization methods in clinical trials, and the framework can be used to study properties of other statistical methods under covariate-adaptive randomization procedures and to define new covariate-adaptive randomization procedures.

The condition (2.6) is important for ensuring that the asymptotic loss of power can be neglected so that the randomization procedure is efficient. When the covariates are discrete, this condition is satisfied for a large family of covariate-adaptive designs, including many popular designs in the literature, and the new family proposed in this paper. However, all covariate-adaptive designs considered in this article are based on discrete covariates. When the covariates are continuous, they are typically discretized in order to be included in the randomization scheme. For-

mula (2.8) gives the cost in terms of power loss of failing to balance the kth covariate completely when a randomization scheme is implemented by breaking down a continuous covariate into a discrete variable with several subcategories. Ma et al. (2024) proposed a procedure that can balance the continuous covariates $\boldsymbol{X}_i$s directly. Let $D_n = \sum_{i=1}^{n}(2T_i - 1)$ and $\boldsymbol{D}_n^x = \sum_{i=1}^{n}(2T_i - 1)\boldsymbol{X}_i$ be defined as in Section 2. Denote the measure of imbalance by $M_n = (D_n)^2 + \|\boldsymbol{D}_n^x\|^2$, $Imb_n^{(1)} = M_n\big|_{T_n=1}$ and $Imb_n^{(2)} = M_n\big|_{T_n=0}$. Ma et al's procedure assigns the nth patient to treatment 1 with probability ρ ($> 1/2$) when $Imb_n^{(1)} < Imb_n^{(2)}$, $1/2$ when $Imb_n^{(1)} = Imb_n^{(2)}$, and $1-\rho$ when $Imb_n^{(1)} > Imb_n^{(2)}$. Ma et al. proved that condition (2.6) is satisfied under some suitable conditions. This procedure can be generalized by defining the allocation probability as $g_n(Imb_n^{(1)} - Imb_n^{(2)})$. We may have similar results to those in Theorems 4-5 and Corollary 3. In particular, if $g_n(x)$ satisfies the conditions (4.5)-(4.7) in Theorem 4, we have

$$LossP_{T,n}(\mu|\boldsymbol{X}) \to 1 \text{ in probability and } \lim_{n\to\infty} SB_n = \frac{1}{2}.$$

The theoretical properties for covariate-adaptive designs can be generalized in several ways. First, the theory and adaptive randomization procedures are based on clinical trials with two treatments, but they can be generalized to multiple treatments (Hu and Zhang, 2004; Tymofyeyev, Rosenberger and Hu, 2007; Hu, Ye and Zhang, 2023). Second, a key as-

sumption to derive theoretical results on the power is the independence between covariates. We may apply a similar idea to dependent covariates by incorporating the correlation structure. Third, to derive the relative loss of power, another important assumption is the normality of the regression errors. It is an open problem whether the conclusions in Theorems 1 and 2 are still valid for a general distribution of errors, especially when an approximate distribution is applied. This may not be an easy problem, but precise self-normalized Cramér-type large deviations are helpful for solving it (c.f. Jing, Shao and Wang (2003)). Fourth, the randomization procedures considered in this paper only depend on the covariates; patient responses to treatments are not incorporated. It is possible to generalize the results to efficient response-adaptive designs (Hu, Zhang and He, 2009) and covariate-adjusted adaptive designs (Zhang et al, 2007). These topics are left for future research.

Supplementary Material

The online supplementary material contains the theoretical proof of all the theorems, and the simulation study of the new CAR with a specific allocation function g in (4.8).

Acknowledgments

The authors thank the Associate Editor and the two anonymous referees for many helpful comments and suggestions. The research was partly supported by grants from NSF of China (No. U23A2064), National Key R&D Program of China (No. 2024YFA1013502) and the Summit Advancement Disciplines of Zhejiang Province (Zhejiang Gongshang University - Statistics).

References

Aickin, M. (2009). A Simulation study of the validity and efficiency of design-adaptive allocation to two groups in the regression situation. *The International Journal of Biostatistics* **5**, 14.

Baldi Antognini, A. (2008). A theoretical analysis of the power of biased coin designs. *Journal of Statistical Planning and Inference* **138**, 1792-1798.

Baldi Antognini, A. and Giovagnoli, A. (2004). A new 'biased coin design' for the sequential allocation of two treatments. *Journal of the Royal Statistical Society. Series C (Applied Statistics)* **53**, 651-664.

Berger, V. W., Ivanova, A. and Knoll, D. M. (2003). Minimizing predictability while retaining balance through the use of less restrictive randomization procedures. *Statistics in medicine*, **22**(19), 3017-3028.

Birkett, N. J. (1985). Adaptive allocation in randomized controlled trials. *Controlled Clinical Trials* **6**, 146-155.

Ciolino, J., Zhao, W., Martin, R., & Palesch, Y (2011). Quantifying the cost in power of ignoring continuous covariate imbalances in clinical trial randomization. *Contemporary Clinical Trials* **32**, 250-259.

Efron, B. (1971). Forcing a sequential experiment to be balanced. *Biometrika* **58**, 403-417.

Feinstein, A. R., and Landis, J. R. (1976). The role of prognostic stratification in preventing the bias permitted by random allocation of treatment. *Journal of Chronic Diseases* **29** 277-284.

Forsythe, A. B. (1987). Validity and power of tests when groups have been balanced for prognostic factors. *Computational Statistics and Data Analysis* **5**, 193-200.

Green, S. B., and Byar, D. P. (1978). The effect of stratified randomization on size and power of statistical tests in clinical trials. *Journal of Chronic Diseases* **31**, 445-454.

Hu, Y. and Hu, F. (2012). Asymptotic properties of covariate-adaptive randomization. *Annals of Statistics* **40**, 1794-1815.

Hu, F. and Rosenberger, W. F. (2006). *The Theory of Response-Adaptive Randomization in Clinical Trials*. John Wiley and Sons. Wiley Series in Probability and Statistics.

Hu, F., Ye, X. and Zhang, L.-X. (2023). Multi-arm covariate-adaptive randomization. *Science China-Mathematics* **66**(1), 163-190.

Hu, F. and Zhang L.-X. (2004). Asymptotic properties of doubly adaptive biased coin designs for multitreatment clinical trials. *The Annals of Statistics* **32**, 268-301.

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

Hu, F., Zhang, L.-X. and He, X. (2009), Efficient randomized adaptive designs. *The Annals of Statistics* **37**, 2543-2560.

Jing, B. Y., Shao, Q. M. and Wang, Q.Y. (2003). Self-normalized Cramér-type large deviations for independent random variables. *Annals of Probability* **31**, 2167-2215.

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

Ma, W., Li, P., Zhang, L.-X. and Hu, F.(2024). A new and unified family of covariate adaptive randomization procedures and their properties. *Journal of the American Statistical Association*, **119**(545), 151-162.

McEntegart, D. J. (2003). The pursuit of balance using stratified and dynamic randomization techniques: An overview. *Drug Information Journal* **37**, 293-308.

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

Rosenberger, W. F. and Lachin, J. M. (2002). *Randomization in Clinical Trials: Theory and Practice*. John Wiley and Sons. Wiley Series in Probability and Statistics.

Rosenberger, W. F. and Sverdlov, O. (2008). Handling covariates in the design of clinical trials. *Statistical Science* **23**, 404-419.

Scott, N. W., McPherson, G. C., Ramsay, C. R. and Campbell, M. K. (2002). The method of minimization for allocation to clinical trials: A review. *Control Clinical Trials* **23**, 662-674.

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

Simon, R. (1979). Restricted aandomization designs in clinical trials. *Biometrics* **35** 503-512.

Smith, R. L. (1984). Properties of biased coin designs in sequential clinical trials. *Annals of Statistics* **12**, 1018-1034.

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

Taves, D. R. (2010). The use of minimization in clinical trials. *Contemporary Clinical Trials* **31**, 180-184.

Tu, D., Shalay, K. and Pater, J. (2000). Adjustment of treatment effect for covariates in clinical trials: statistical and regulatory issues. *Drug Information Journal* **31**, 180-184.

Tymofyeyev, Y., Rosenberger, W.F. and Hu, F. (2007). Implementing optimal allocation in sequential binary response experiments. *Journal of the American Statistical Association* **102**, 224-234.

Wei, L. J. (1978). The adaptive biased coin design for sequential experiments. *Annals of Statistics* **6**, 92- 100

Zhang, L.X., Hu, F., Cheung. S.H. and Chan, W.S. (2007). Asymptotic properties of covariate-

adjusted adaptive designs. *Annals of Statistics* **35**, 1166-1182.

Zhu, H., Zhang, L.-X., Ning, J. and Wang, L. (2025). Testing hypotheses of covariate-adaptive randomized clinical trials with time-to-event outcomes. *Statistica Sinica* **35**, 1051-1067.

Li-Xin Zhang

School of Statistics and Data Sciences, Zhejiang Gongshang University, 310018

E-mail: stazlx@mail.zjgsu.edu.cn

ORCID ID: https://orcid.org/0000-0002-0121-0187