Journal of the
American
Statistical
Association

Journal of the American Statistical Association

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/uasa20

A New and Unified Family of Covariate Adaptive Randomization Procedures and Their Properties

Wei Ma, Ping Li, Li-Xin Zhang & Feifang Hu

To cite this article: Wei Ma, Ping Li, Li-Xin Zhang & Feifang Hu (2022): A New and Unified Family of Covariate Adaptive Randomization Procedures and Their Properties, Journal of the American Statistical Association, DOI: 10.1080/01621459.2022.2102986

To link to this article: https://doi.org/10.1080/01621459.2022.2102986

JOURNAL OF THE AMERICAN STATISTICAL ASSOCIATION
2022, VOL. 00, NO. 0, 1–12: Theory and Methods
https://doi.org/10.1080/01621459.2022.2102986

Taylor & Francis
Taylor & Francis Group

A New and Unified Family of Covariate Adaptive Randomization Procedures and Their Properties

Wei Ma[a], Ping Li[b], Li-Xin Zhang[c], and Feifang Hu[d]

[a]Institute of Statistics and Big Data, Renmin University of China, Beijing, China; [b]LinkedIn Corporation, Bellevue, WA; [c]Center for Data Science and School of Mathematical Sciences, Zhejiang University, Hangzhou, China; [d]Department of Statistics, George Washington University, Washington, DC

ABSTRACT

In clinical trials and other comparative studies, covariate balance is crucial for credible and efficient assessment of treatment effects. Covariate adaptive randomization (CAR) procedures are extensively used to reduce the likelihood of covariate imbalances occurring. In the literature, most studies have focused on balancing of discrete covariates. Applications of CAR with continuous covariates remain rare, especially when the interest goes beyond balancing only the first moment. In this article, we propose a family of CAR procedures that can balance general covariate features, such as quadratic and interaction terms. Our framework not only unifies many existing methods, but also introduces a much broader class of new and useful CAR procedures. We show that the proposed procedures have superior balancing properties; in particular, the convergence rate of imbalance vectors is $O_P(n^{\epsilon})$ for any $\epsilon > 0$ if all of the moments are finite for the covariate features, relative to $O_P(\sqrt{n})$ under complete randomization, where n is the sample size. Both the resulting convergence rate and its proof are novel. These favorable balancing properties lead to increased precision of treatment effect estimation in the presence of nonlinear covariate effects. The framework is applied to balance covariate means and covariance matrices simultaneously. Simulation and empirical studies demonstrate the excellent and robust performance of the proposed procedures. Supplementary materials for this article are available online.

ARTICLE HISTORY
Received August 2021
Accepted July 2022

KEYWORDS
Covariate adaptive randomization; Covariate balance; Imbalance vector; Markov chain; Treatment effect estimation

1. Introduction

Balancing of baseline covariates is crucial in clinical trials and other comparative studies, such as online A/B tests and economic field experiments. The main purposes are to enhance the credibility of trial results and to increase the efficiency of treatment effect estimation. Covariate adaptive randomization (CAR) procedures have often been used to achieve these goals (McEntegart 2003; Taves 2010; Lin, Zhu, and Su 2015). Many such procedures have been proposed to balance treatment assignments within strata and over covariate margins, provided that the covariates under consideration are discrete (categorical with two or more levels) or have been discretized (Pocock and Simon 1975; Baldi Antognini and Zagoraiou 2011; Hu and Hu 2012; Rosenberger and Lachin 2015). In contrast, studies using CAR procedures with continuous covariates have been comparatively rare, despite the availability of various empirical methods (some of which are reviewed below). This article describes a unified family of CAR procedures that are applicable to discrete or continuous covariates or their combinations. These novel procedures are tractable and have favorable properties for covariate balancing and treatment effect estimation.

In CAR procedures, balancing of continuous covariates is achieved by minimization of prescribed imbalance measures. The simplest and most intuitive approaches target differences in covariate means (Li, Zhou, and Hu 2019) or related parameters, such as Mahalanobis distances (Qin et al. 2018; Zhou et al. 2018) and the p-values of analysis of variance (Frane 1998). These methods are appealing because covariate means are routinely reported as indicators of how well baseline covariates are balanced, but their limitation is that higher-order moments and other important covariate features are neglected. Consequently, a variety of methods have attempted to balance covariates based on additional features, such as variance (Nishi and Takaichi 2004; Endo et al. 2006), rank (Hoehler 1987; Stigsby and Taves 2010), and certain nonparametric estimators of covariate distributions (Lin and Su 2012; Ma and Hu 2013; Jiang, Ma, and Yin 2018); see Hu et al. (2014) for a comprehensive review. However, these methods have been created for ad hoc purposes, and their theoretical properties remain largely unknown, limiting their applicability. Model-based approaches have also been proposed to improve efficiency using optimal design theory (Begg and Iglewicz 1980; Atkinson 1982; Smith 1984; Begg and Kalish 1984; Atkinson 2002). By assuming a linear model between the responses and covariates, Atkinson's D_A-optimal biased coin design (Atkinson 1982) achieves a balanced allocation over the covariates; in general, however, model-based approaches may not imply balance (Rosenberger and Lachin 2015).

To assess how covariates are balanced, and thus assure the validity of a CAR procedure, the convergence rate of the

CONTACT Feifang Hu ✉ feifang@gwu.edu ✉ Department of Statistics, George Washington University, Washington, DC.

Supplementary materials for this article are available online. Please go to www.tandfonline.com/r/JASA.

imbalance vectors must be described. The imbalance vectors typically have the form of $\sum_{i=1}^{n}(2T_i - 1)X_i$, where n is the sample size and $T_i = 1$ for the treatment and $T_i = 0$ for the control. When X_i is the indicator of a covariate margin or stratum, the form reduces to marginal or within-stratum imbalances, which have been well studied for various CAR procedures with discrete covariates (e.g., Baldi Antognini and Zagoraiou 2011; Hu and Hu 2012; Ma, Hu, and Zhang 2015; Rosenberger and Lachin 2015). The best available convergence rate of the imbalance vectors is $O_P(1)$, relative to $O_P(\sqrt{n})$ under complete randomization (Hu and Hu 2012; Ma, Hu, and Zhang 2015). Recently, $O_P(1)$ rates have been obtained under certain scenarios with continuous covariate vectors X_i (Qin et al. 2018; Li, Zhou, and Hu 2019). To ensure that the Markov chains induced by the imbalance vectors are irreducible (Meyn and Tweedie 2009), these studies made strong implicit assumptions about the covariates; for example, that the multivariate density functions were positive everywhere. However, suppose the interest is to consider a more general form of imbalance vector, $\sum_{i=1}^{n}(2T_i - 1)\phi(X_i)$, where $\phi(X_i)$ contains additional covariate features or combinations of both discrete and continuous covariates. In such cases, the irreducibility requirement is generally not met or is at least difficult to verify, presenting a challenge to the balancing of covariates beyond the first moment. However, we find that we can obtain a slightly weaker conclusion (but making almost no difference from a practical point of view) using only moment conditions: the convergence rate of the imbalance vectors is $O_P(n^{\epsilon})$ for any $\epsilon > 0$ if all of the moments of $\phi(X_i)$ are finite.

It is generally accepted that more balanced allocation over covariates is associated with higher statistical efficiency. This assertion has been confirmed for CAR procedures with discrete covariates by several recent works (Shao, Yu, and Zhong 2010; Ma, Hu, and Zhang 2015; Bugni, Canay, and Shaikh 2018; Ma, Tu, and Liu 2020b). Moreover, when a balance of continuous covariates is pursued, and only the first-order linear covariate effect is assumed to be present, Ma et al. (2020a) showed that the balancing of covariate means can increase the precision of treatment effect estimation. Any convergence rate of $o_P(\sqrt{n})$ for the imbalance vector $\sum_{i=1}^{n}(2T_i - 1)X_i$ suffices to guarantee optimal precision under certain assumptions. However, nonlinear covariate effects are common in many applications. From the statistical inference point of view, it is of interest to consider the balancing of covariates beyond the first moment. To ground our discussion, we examine a randomized clinical trial of depression (Keller et al. 2000). A post hoc analysis of the trial data clearly demonstrates a nonlinear quadratic age effect, and suggests that covariate balancing for these data would ideally address the second moment and other nonlinear features of the covariates (see Section 6 for details). A favorable convergence rate of the imbalance vectors, $\sum_{i=1}^{n}(2T_i - 1)\phi(X_i)$, is essential to achieve a more accurately estimated treatment effect in the presence of nonlinear covariate effects.

In this article, we propose a family of CAR procedures to improve the balancing of general covariate features. The proposed procedures use an adaptive randomization scheme to sequentially minimize the imbalance measure associated with the feature map $\phi(X_i)$. This framework unifies many recently proposed methods and, more importantly, can generate a broader range of CAR procedures because of its flexibility in defining various feature maps $\phi(X_i)$. Based on this framework, we develop a new CAR procedure that can balance covariate means and covariance matrices between treatments. Moreover, we obtain the convergence rate of covariate imbalance vectors for the proposed procedures with finite-dimensional feature maps $\phi(X_i)$. The proof relies only on moment conditions, without requiring irreducibility for the Markov chain induced by the imbalance vector. Both the convergence rate results and the proof techniques are new to the discipline of CAR. In addition, under further assumptions that ensure that the Markov chain is irreducible, stronger and more complete results can be obtained. Finally, under both an additive treatment effect model and a more general outcome model, we derive the asymptotic behaviors of the difference-in-means estimator for the treatment effect. These asymptotic results show that when nonlinear covariate effects are present, the proposed procedures lead to more precise treatment effect estimation.

The remainder of this article is organized as follows. In Section 2, we describe the proposed procedures. In Section 3, we present the theoretical properties. Treatment effect estimation is discussed in Section 4. Simulation studies and a clinical trial example are presented in Sections 5 and 6. Section 7 summarizes the study and provides directions for future work. Technical proofs and additional simulations are provided in the Supplementary Appendix.

2. A New and Unified Family of CAR Procedures

2.1. General Framework

Suppose that n units are to be assigned to two treatment groups. Let T_i be the assignment of the ith unit, such that $T_i = 1$ for the treatment and $T_i = 0$ for the control. Let $n_1 = \sum_{i=1}^{n} T_i$ and $n_0 = \sum_{i=1}^{n}(1 - T_i)$ denote the numbers of treated units and control units, respectively. Denote by $X_i = (x_{i1}, \ldots, x_{ip})^{\mathrm{T}}$ a p-dimensional vector of baseline covariates for the ith unit.

We propose to balance general covariate features $\phi(X_i)$, defined by a feature map $\phi(X_i) : \mathbb{R}^p \mapsto \mathbb{R}^q$ that maps X_i into a q-dimensional feature space. Here, q is usually larger than p so that $\phi(X_i)$ has more features than the original covariates. We define the imbalance measure Imb_n as the squared Euclidean norm of the imbalance vector $\sum_{i=1}^{n}(2T_i - 1)\phi(X_i)$,

$$\mathrm{Imb}_n = \left\| \sum_{i=1}^{n}(2T_i - 1)\phi(X_i) \right\|^2. \quad (1)$$

The proposed procedure to minimize the imbalance measure Imb_n is defined as follows:

(a) Randomly assign the first unit with equal probability to the treatment or to the control.
(b) Suppose $(n - 1)$ units have been assigned to a treatment $(n > 1)$ and the nth unit is to be assigned. Calculate the "potential" imbalance measures $\mathrm{Imb}_n^{(1)}$ and $\mathrm{Imb}_n^{(0)}$, corresponding to $T_n = 1$ and $T_n = 0$, respectively.

(c) Assign the nth unit to the treatment with the probability

$$P(T_n = 1 | X_n, \ldots, X_1, T_{n-1}, \ldots, T_1) = \begin{cases} \rho & \text{if } \mathrm{Imb}_n^{(1)} < \mathrm{Imb}_n^{(0)}, \\ 1-\rho & \text{if } \mathrm{Imb}_n^{(1)} > \mathrm{Imb}_n^{(0)}, \\ 0.5 & \text{if } \mathrm{Imb}_n^{(1)} = \mathrm{Imb}_n^{(0)}, \end{cases}$$

where $0.5 < \rho \leq 1$. Note that $\mathrm{Imb}_n^{(1)} - \mathrm{Imb}_n^{(0)} = 4\{\sum_{i=1}^{n-1}(2T_i - 1)\phi(X_i)\}^{\mathrm{T}}\phi(X_n)$.

(d) Repeat the last two steps until all units are assigned.

Remark 2.1. It is suggested that larger values of the biasing probability ρ, such as 0.85, 0.90, and 0.95, should be used when covariates are involved (Hu and Hu 2012). The value of ρ is set to 0.90 throughout this article. In addition, more general allocation functions could be used instead of Efron's biased coin function (Efron 1971; Hu and Zhang 2020).

The procedure is flexible in application because various approaches can be used to define the feature map and the corresponding imbalance measure. The following are some examples of the imbalance measures used in existing CAR procedures.

Example 2.1 (Covariate means). Defining $\phi(X_i) = X_i$, the imbalance measure is

$$\mathrm{Imb}_n = \left\| \sum_{i=1}^{n} (2T_i - 1) X_i \right\|^2 = \Bigl(\sum_{i:T_i=1} X_i - \sum_{i:T_i=0} X_i \Bigr)^{\mathrm{T}} \Bigl(\sum_{i:T_i=1} X_i - \sum_{i:T_i=0} X_i \Bigr).$$

This imbalance measure is proportional to the squared difference of the sample means, provided each treatment group has the same number of units. The univariate case of an imbalance measure of this type is considered in Li, Zhou, and Hu (2019), with a restriction imposed to ensure that the two treatments are assigned approximately equal numbers of units within a prespecified tolerance.

Example 2.2 (Mahalanobis distance). The Mahalanobis distance of the covariate means has been used in the literature as an imbalance measure (Morgan and Rubin 2012; Qin et al. 2018; Zhou et al. 2018). Suppose the covariance matrix $\mathrm{cov}(X_i)$ is known, and the eigendecomposition is $\mathrm{cov}(X_i)^{-1} = VDV^{\mathrm{T}}$. Letting $\phi(X_i) = D^{1/2}V^{\mathrm{T}}X_i$, the imbalance measure is

$$\mathrm{Imb}_n = \left\| \sum_{i=1}^{n} (2T_i - 1) D^{1/2} V^{\mathrm{T}} X_i \right\|^2 = \Bigl(\sum_{i:T_i=1} X_i - \sum_{i:T_i=0} X_i \Bigr)^{\mathrm{T}} \mathrm{cov}(X_i)^{-1} \Bigl(\sum_{i:T_i=1} X_i - \sum_{i:T_i=0} X_i \Bigr),$$

which is proportional to the Mahalanobis distance if each treatment has the same number of units. Sequential rerandomization and pairwise allocation can be used, as in Zhou et al. (2018) and Qin et al. (2018), respectively, to ensure that the two treatments are each assigned the same number of units.

Example 2.3 (Discrete covariates). Hu and Hu (2012) considered an imbalance measure that simultaneously accounts for overall, marginal, and within-stratum imbalances, denoted by D_n, $D_n(l; k_l^*)$, and $D_n(k_1^*, \ldots, k_p^*)$, respectively. Let l denote the lth covariate within a subject containing p covariates. The imbalance measure in Hu and Hu (2012)

$$\mathrm{Imb}_n = w_o D_n^2 + \sum_{l=1}^{p} w_{m,l}\{D_n(l; k_l^*)\}^2 + w_s\{D_n(k_1^*, \ldots, k_p^*)\}^2$$

can be obtained by adopting the following feature map $\phi(X_i)$ for discrete covariates X_i,

$$\phi = (\sqrt{w_o}, \underbrace{\ldots, \sqrt{w_{m,l}}\delta_{(l;k_l^*)}, \ldots,}_{\sum m_l \text{ marginal terms}} \underbrace{\ldots, \sqrt{w_s}\delta_{(k_1^*,\ldots,k_p^*)}, \ldots}_{\prod m_l \text{ within-stratum terms}})^{\mathrm{T}},$$

where $\delta_{(l;k_l^*)}$ is the indication function for the lth covariate belonging to the margin $(l; k_l^*)$, and $\delta_{(k_1^*,\ldots,k_p^*)}$ is the indication function for a covariate belonging to the stratum $(k_1^*, \ldots, k_p^*)$, $1 \leq l \leq p$ and $1 \leq k_l^* \leq m_l$. w_o, $w_{m,l}$, and w_s are nonnegative weights applied to the overall, marginal, and within-stratum imbalances, respectively. By choosing different weights, the method of Hu and Hu (2012) can include several important CAR procedures as special cases, such as minimization (Pocock and Simon 1975), if $w_o = w_s = 0$, and stratified biased coin design (Shao, Yu, and Zhong 2010), if $w_o = w_{m,l} = 0$.

2.2. Balancing Covariate Means and Covariance Matrices

Suppose the covariate sample mean and the covariance matrix are $\bar{X}_a = \sum_{i:T_i=a} X_i / n_a$ and $S_a = \sum_{i:T_i=a} X_i X_i^{\mathrm{T}} / n_a$, respectively, under the treatment ($a = 1$) and control ($a = 0$), and we intend to minimize

$$w_1 ||\bar{X}_1 - \bar{X}_0||^2 + w_2 ||S_1 - S_0||_{\mathrm{F}}^2, \tag{2}$$

where $||\cdot||_{\mathrm{F}}$ denotes the Frobenius norm. w_1 and w_2 are nonnegative weights applied to the mean and covariance imbalances, respectively. To reduce the imbalance measure presented in (2), we next introduce a new procedure, termed COV, belonging to the CAR family proposed in Section 2.1. We observe that the quantity in (2) is equal to $w_1(\bar{X}_1 - \bar{X}_0)^{\mathrm{T}}(\bar{X}_1 - \bar{X}_0) + w_2 \mathrm{vec}\,(S_1 - S_0)^{\mathrm{T}}\,\mathrm{vec}\,(S_1 - S_0)$, where $\mathrm{vec}(\cdot)$ converts a matrix into a column vector.

The COV Procedure: A CAR procedure of the form proposed in Section 2.1 with its imbalance measure defined by the feature map,

$$\phi(X_i) = (\sqrt{w_0}, \sqrt{w_1}X_i^{\mathrm{T}}, \sqrt{w_2}\mathrm{vec}(X_i X_i^{\mathrm{T}})^{\mathrm{T}})^{\mathrm{T}}, \quad w_0, w_1, w_2 \geq 0.$$

Here, we consider an imbalance measure designed to balance means and covariance matrices simultaneously. In contrast to imbalance measures with potentially even higher-dimensional feature maps, COV focuses on the first two moments of the covariates because, in many applications, the quadratic effects and two-factor interactions are more relevant than higher-order effects. The weight w_0 and the constant function with a value of 1 are added into the feature map to control the overall difference $\sum_{i=1}^{n}(2T_i - 1) = n_1 - n_0$. By this approach, we can directly control the overall difference via sequential allocation without requiring pairwise allocation or imposing additional restrictions on the randomization.

There is a tradeoff between the mean and covariance imbalances in the imbalance measure. A larger weight placed on the mean imbalance will increase its role in determining the allocation of the next unit, and vice versa. For the p-dimensional covariate X_i, there are p^2 terms in the covariance matrix and only p terms in the mean vector; imposing equal weights will cause the imbalance measure to be dominated by the covariance imbalances. Unless otherwise stated, our simulations use $w_1/w_2 = p$ to make the magnitudes of the mean and covariance imbalances more comparable.

Remark 2.2. If $w_1 > 0$ and $w_0 = w_2 = 0$, the COV procedure reduces to a variant of the method of Qin et al. (2018), which allocates units sequentially if the covariates are available and standardized prior to randomization. Although it can effectively minimize the covariate mean imbalance, this method may incur unsatisfactorily large covariance matrix imbalances. In contrast, the unreduced COV procedure trades off some imbalance in covariate means to address covariance matrix imbalances and improve overall covariate balance.

Remark 2.3. Suppose the cross-product terms in the covariance matrix are of no interest. For example, if the interaction effects of the covariates are known to be weak and can therefore be ignored, then variances of each covariate can be balanced without consideration of their correlations. In such cases, the COV imbalance measure reduces to a form similar to that of Nishi and Takaichi (2004). However, the proposed procedure is more general and adds supporting theoretical justifications.

2.3. *Extension to Kernelized Imbalance Measure*

So far, all of the discussed examples have constructed feature maps $\phi(X_i)$ explicitly. Using the "kernel trick" in machine learning, we can define the imbalance measure using a kernel function $k(\cdot,\cdot)$ alone and construct feature maps implicitly. Defining $k(X_i, X_j) = \phi(X_i)^{\mathrm{T}}\phi(X_j)$, the imbalance measure (1) can be rewritten as

$$\mathrm{Imb}_n = \sum_{i,j}(2T_i - 1)(2T_j - 1)k(X_i, X_j),$$

where i and j range from 1 to n. This form can be applied with a wide variety of machine learning kernels, such as polynomial, spline, and ANOVA kernels (Hofmann, Schölkopf, and Smola 2008), to constitute a broad family of new CAR procedures. However, these methods may be less intuitive to interpret than cases with user-specified feature maps. We emphasize that the kernel method presented here is distinct from some previous CAR procedures that have used related kernel smoothing techniques, such as kernel density estimation, that can apply the same kernel functions (Ma and Hu 2013; Jiang, Ma, and Yin 2018).

In this article, we adopt the Gaussian kernel because it is commonly used and has favorable properties; for example, it has been shown to be a universal and characteristic kernel (Sriperumbudur, Fukumizu, and Lanckriet 2011).

The KER Procedure: A CAR procedure of the form proposed in Section 2.1 with its imbalance measure defined by the Gaussian kernel function,

$$k(X_i, X_j) = \exp\left(-\frac{||X_i - X_j||^2}{2\sigma^2}\right).$$

The explicit form of the feature map, $\phi(X_i)$, corresponding to the Gaussian kernel can be derived as follows (Steinwart, Hush, and Scovel 2006). Assuming $X_i \in \mathbb{R}^1$ and letting $\sigma^2 = 1/2$ for convenience, the basis functions of the feature map are given by

$$\phi(X_i) = \exp(-X_i^2)\left(1, \sqrt{\frac{2}{1!}}X_i, \sqrt{\frac{2^2}{2!}}X_i^2, \ldots\right).$$

For the general case $X_i \in \mathbb{R}^p$, the basis functions are the tensor products of the basis functions given above. In particular, the first basis function of the Gaussian kernel is $\exp(-||X_i||^2)$.

Remark 2.4. A positive definite kernel function $k(\cdot,\cdot)$ defines a reproducing kernel Hilbert space (RKHS) $\mathcal{H}$, with inner product $\langle\cdot,\cdot\rangle_{\mathcal{H}}$ and norm $||\cdot||_{\mathcal{H}}$ (Aronszajn 1950). The proposed imbalance measure is the squared norm of the imbalance vector, $\sum_{i=1}^{n}(2T_i - 1)k(X_i,\cdot)$, in the RKHS $\mathcal{H}$, giving $\mathrm{Imb}_n = ||\sum_{i=1}^{n}(2T_i - 1)k(X_i,\cdot)||^2_{\mathcal{H}}$. Therefore, the proposed CAR procedure balances the covariates in a kernel-induced RKHS.

Remark 2.5. The imbalance measure defined by a characteristic kernel, such as the Gaussian kernel, is closely associated with the concept of maximum mean discrepancy (MMD), which is a probability distribution metric based on distances between kernel mean embeddings (Gretton et al. 2008; Muandet et al. 2016). Under equal sample size, Imb_n is proportional to the square of an empirical estimate of MMD (Gretton et al. 2008, p. 6). This observation offers another interpretation of the proposed CAR procedure (with a characteristic kernel): it attempts to improve the overall distributional similarity of the covariates by sequentially minimizing a probability distribution metric (the MMD) between different treatments.

3. Theoretical Properties

3.1. *General Results*

Consider the proposed CAR procedure with the finite-dimensional feature map, $\phi(X) = (\phi_1(X), \ldots, \phi_q(X))^{\mathrm{T}} : \mathbb{R}^p \mapsto \mathbb{R}^q$. To obtain the convergence rate of the imbalance vector, $\Lambda_n = \sum_{i=1}^{n}(2T_i - 1)\phi(X_i)$, we require the following assumptions:

Assumption 1. The covariates $\{X_i = (x_{i1}, \ldots, x_{ip})^{\mathrm{T}}\}_{i=1}^{n}$ are independent copies of X.

Assumption 2. The feature map $\phi(X)$ satisfies that $E\{||\phi(X)||^{\gamma}\}$ is finite for a given $\gamma > 2$.

Assumption 1 ensures that $\{\Lambda_n\}$ is a Markov chain on $\mathbb{R}^q$ under the proposed CAR procedure. To see this, note that $\Lambda_{n+1} = \Lambda_n - \Delta_{n+1}$, with $\Delta_{n+1} = (-1)^{T_{n+1}}\phi(X_{n+1})$. Given Λ_n, Δ_{n+1} is conditionally independent of $\{\Lambda_1, \ldots, \Lambda_{n-1}\}$, and thus Λ_{n+1} is also conditionally independent of $\{\Lambda_1, \ldots, \Lambda_{n-1}\}$.

Assumption 2 concerns only the moment condition of the covariate features. Here, we do not assume any special properties of the Markov chain $\{\Lambda_n\}$, such as irreducibility and aperiodicity.

Theorem 3.1. Suppose Assumptions 1 and 2 hold. Then $E(||\Lambda_n||^2) = O(n^{\frac{1}{\gamma-1}})$ and so, $\Lambda_n = O_P(n^{\frac{1}{2(\gamma-1)}}) = o_P(\sqrt{n})$. In particular, if $E\{||\phi(X)||^\gamma\}$ is finite for all $\gamma > 2$, then $\Lambda_n = O_P(n^\epsilon)$ for any $\epsilon > 0$.

The results imply that $\sum_{i=1}^n (2T_i - 1)\phi_j(X_i)$ is also $O_P(n^{\frac{1}{2(\gamma-1)}})$ for each $j = 1, \ldots, q$. Note that the convergence rate is $O_P(n^{1/6})$ if $\phi(X)$ has a finite fourth moment. This convergence rate is better than the rate of $O_P(\sqrt{n})$ under complete randomization and many existing CAR procedures (e.g., Atkinson 1982; Jiang, Ma, and Yin 2018). Furthermore, the rate is $O_P(n^\epsilon)$ for any $\epsilon > 0$ if all of the moments of $\phi(X)$ are finite. This condition is satisfied, for example, when $\phi(X)$ is bounded or comprises polynomials of normally distributed covariates.

Remark 3.1. The rate of $O_P(n^\epsilon)$ for any $\epsilon > 0$ is not necessarily sharp. The rate of $O_P(1)$ is obtained for some important special cases of the proposed procedures (Hu and Hu 2012; Qin et al. 2018; Li, Zhou, and Hu 2019). However, these results require additional assumptions (e.g., that $\phi(X)$ is discrete or has a continuous positive density) to ensure that the Markov chain $\{\Lambda_n\}$ is irreducible. A general condition to ensure the irreducibility is presented in Section 3.2 for the case of continuous covariates. However, these assumptions may fail for the COV procedure with $w_0 > 0$, or more generally, when the procedure is intended to balance mixed (both discrete and continuous) covariate profiles; see Section B.4 in the Supplementary Appendix for an example. In contrast, the results presented in Theorem 3.1 only require moment conditions and are thus more general and applicable to both discrete and continuous covariates and their combinations. Although the derived rate is slightly slower than $O_P(1)$, the difference is negligible for sample sizes typically encountered in practice. For example, Phase III clinical trials often enroll hundreds of patients.

The following corollary is an immediate application of Theorem 3.1 to the COV procedure.

Corollary 3.2. Suppose Assumptions 1 and 2 hold for the covariates $\{X_i\}_{i=1}^n$ and the feature map $\phi(X_i) = (\sqrt{w_0}, \sqrt{w_1}X_i^{\mathrm{T}}, \sqrt{w_2}\mathrm{vec}(X_iX_i^{\mathrm{T}})^{\mathrm{T}})^{\mathrm{T}}$ with $w_0 > 0, w_1 > 0$, and $w_2 > 0$. Then $\Lambda_n = o_P(\sqrt{n})$. In particular,

(i) $\sum_{i=1}^n (2T_i - 1) = n_1 - n_0 = o_P(\sqrt{n})$;
(ii) $\sum_{i=1}^n (2T_i - 1)x_{ij} = o_P(\sqrt{n})$ for any $j = 1, \ldots, p$;
(iii) $\sum_{i=1}^n (2T_i - 1)x_{ij}x_{ij'} = o_P(\sqrt{n})$ for any $j, j' = 1, \ldots, p$.

Furthermore, if $E(|x_{ij}|^\gamma) < \infty$ for all $\gamma > 2$ and $j = 1, \ldots, p$, then the above statements still hold with $o_P(\sqrt{n})$ replaced by $O_P(n^\epsilon)$ for any $\epsilon > 0$,.

The conclusions in Theorem 3.1 require that the feature map $\phi(X)$ is finite-dimensional. Achieving similar results for infinite-dimensional $\phi(X)$ needs an additional assumption, which, roughly speaking, requires that the intrinsic dimension of $\phi(X)$ is finite.

Assumption 3. The feature map $\phi(X) = (\phi_1(X), \phi_2(X), \ldots)^{\mathrm{T}}$ satisfies requirements that the infinite-dimensional matrix $\Sigma = E\{\phi(X)\phi(X)^{\mathrm{T}}\} = \big(E\{\phi_i(X)\phi_j(X)\} : i, j = 1, 2, \ldots\big)$ has only a finite number of nonzero eigenvalues $\lambda_1, \ldots, \lambda_{d'}$, and that there exists an orthogonal matrix U such that $U\Sigma U^{\mathrm{T}} = \mathrm{diag}(\lambda_1, \ldots, \lambda_{d'}, 0, \ldots)$.

Theorem 3.3. Suppose Assumptions 1–3 hold. Then $E(||\Lambda_n||^2) = O(n^{\frac{1}{\gamma-1}})$ and so, $\Lambda_n = O_P(n^{\frac{1}{2(\gamma-1)}}) = o_P(\sqrt{n})$.

Remark 3.2. The above results may not hold for truly infinite-dimensional feature maps, such as those using the Gaussian kernel in the KER procedure. The proofs of Theorems 3.1 and 3.3 do not follow through because the key inequality (A.2) in the Supplementary Appendix may not be true for the infinite-dimensional case. The simulation results in Table 1 and Table B.1 in the Supplementary Appendix indicate that the first few bases are balanced quite well under KER, but the imbalances on later bases have a clear increasing trend with sample size. However, the imbalances are still smaller than those obtained under complete randomization. These observations illustrate the challenges of balancing high-dimensional covariate features in the design stage. Further investigation will be needed to accurately describe the stochastic behavior of the imbalance vectors induced by infinite-dimensional feature maps, which is important to fully characterize the balancing properties of the proposed procedures.

3.2. Further Results

To consider treatment effect estimation under a general outcome model in Section 4.2, we provide further results on the properties of the proposed procedure. These results require the following assumption that, together with Assumptions 1 and 2, ensure that $\{\Lambda_n\}$ is ψ-irreducible (Meyn and Tweedie 2009):

Assumption 4. Suppose that the distribution of $\phi(X)$ is Γ_ϕ, and there is an n_c and a constant $1 \geq c_\nu > 0$ such that

$$\Gamma_\phi^{n_c*}(A) \geq c_\nu \int_A \nu(y)dy \text{ for any Borel set } A,$$

where $\nu(y)$ is a density function with $\inf_{y \in O} \nu(y) > 0$ for an open set O, and Γ_ϕ^{k*} is the kth convolution of Γ_ϕ.

Theorem 3.4. Suppose Assumptions 1, 2 (with $\gamma = 2$) and 4 hold. Then $\{\Lambda_n\}$ is a positive Harris recurrent Markov chain and $\Lambda_n = O_P(1)$.

Theorem 3.5. Suppose Assumptions 1, 2 (with $\gamma > 5$) and 4 hold. Assume that $m(X)$ is a function of X with $E\{|m(X)|\} < \infty$. Then

$$\frac{1}{n}\sum_{i=1}^n (2T_i - 1)m(X_i) \xrightarrow{P} 0.$$

Table 1. Standard deviations of imbalance vectors under various randomization procedures.

Randomization	n	$\sum(2T_i-1)$			$\sum(2T_i-1)x_{i,1}$			$\sum(2T_i-1)x_{i,1}^2$			$\sum(2T_i-1)e^{-\|X_i\|^2}$		
		$p=1$	$p=2$	$p=4$	$p=1$	$p=2$	$p=4$	$p=1$	$p=2$	$p=4$	$p=1$	$p=2$	$p=4$
CR	200	14.22	13.95	14.13	13.84	14.38	14.13	24.75	24.39	24.75	9.54	6.28	2.81
	500	22.11	22.77	22.53	22.24	22.20	22.46	38.47	39.11	39.24	14.99	10.18	4.55
	800	28.39	28.57	27.84	28.39	27.71	28.64	48.91	48.77	49.26	18.91	12.64	5.66
	2000	44.24	44.83	44.78	45.20	44.93	43.78	75.80	76.94	77.30	29.89	20.08	8.86
COV	200	14.01	14.06	14.08	1.04	1.29	1.69	24.21	24.43	24.32	9.43	6.32	2.80
$(w_0,w_1,w_2)=(0,1,0)$	500	22.17	22.72	22.30	1.06	1.29	1.66	38.84	39.24	38.74	14.88	10.12	4.47
	800	28.04	28.34	28.80	1.03	1.28	1.65	48.66	48.73	49.74	18.77	12.83	5.66
	2000	43.31	44.75	44.51	1.05	1.28	1.69	75.31	78.17	77.42	29.38	20.08	8.89
COV	200	1.01	1.30	1.68	1.36	1.53	1.88	21.02	20.23	20.09	5.00	4.41	2.46
$(w_0,w_1,w_2)=(1,1,0)$	500	1.04	1.30	1.67	1.34	1.52	1.85	32.62	32.24	32.39	7.91	6.91	3.90
	800	0.97	1.27	1.69	1.34	1.52	1.85	41.12	41.46	39.88	9.82	8.72	4.79
	2000	1.00	1.29	1.70	1.35	1.52	1.86	65.09	63.95	62.73	15.54	13.69	7.67
COV	200	14.51	11.41	9.00	2.04	2.20	2.58	2.91	3.95	6.09	11.01	6.71	2.92
$(w_0,w_1,w_2)=(0,p,1)$	500	22.79	18.40	14.38	2.09	2.24	2.66	2.99	3.83	6.14	17.38	10.76	4.61
	800	29.03	23.08	18.01	2.10	2.15	2.63	3.00	3.97	6.16	22.09	13.68	5.83
	2000	44.91	36.91	28.27	2.07	2.23	2.64	2.93	3.96	6.08	34.09	21.55	9.18
COV	200	1.47	2.32	4.28	2.29	2.35	2.65	3.01	4.02	6.08	3.75	3.57	2.03
$(w_0,w_1,w_2)=(1,p,1)$	500	1.44	2.30	4.33	2.25	2.30	2.68	3.12	4.00	6.14	5.86	5.84	3.41
	800	1.49	2.34	4.34	2.30	2.28	2.66	3.04	4.03	6.21	7.39	7.29	4.36
	2000	1.50	2.34	4.45	2.29	2.28	2.62	3.14	4.06	6.16	11.68	11.82	7.29
KER	200	1.56	2.58	4.94	3.26	4.55	7.06	8.84	11.19	15.31	0.79	0.78	0.70
	500	1.65	2.79	6.08	3.55	5.38	9.10	10.67	14.00	20.40	0.80	0.79	0.75
	800	1.66	3.00	6.77	3.74	5.69	10.29	11.79	15.56	23.72	0.79	0.80	0.76
	2000	1.71	3.20	7.68	4.16	6.54	12.71	13.53	19.05	30.76	0.80	0.78	0.77

Theorem 3.6. Suppose Assumptions 1, 2 (with $\gamma > 5$) and 4 hold. Assume that $m(X)$ is a function of X with $E\{m^2(X)\} < \infty$. Then there is a $\sigma_m \geq 0$ such that

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n}(2T_i-1)m(X_i) \xrightarrow{D} N(0,\sigma_m^2).$$

Theorems 3.5 and 3.6 provide a novel law of large numbers and a central limit theorem for the proposed CAR procedure under assumptions that ensure that $\{\Lambda_n\}$ is ψ-irreducible. In Theorem 3.6, $\sigma_m^2 = 0$ if $m(X) \in \text{span}\{\phi(X)\} = \{\beta^{\mathrm{T}}\phi(X) | \beta \in \mathbb{R}^q\}$. In general, σ_m^2 depends on the invariant probability measure of $\{\Lambda_n\}$ and may not have a closed form.

Remark 3.3. Assumption 4 is obviously satisfied if $\phi(X)$ has a continuous positive density on an open set ($n_c = 1$). More generally, Assumption 4 is satisfied if the sum of a finite number of independent copies of $\phi(X)$ has a continuous positive density on an open set, in which case we call the distribution of $\phi(X)$ spread out (Meyn and Tweedie 2009, p. 107). The assumption is mild, as illustrated by the below example balancing for multiple moments.

Example 3.1 (Balancing higher moments). Suppose that the one-dimensional covariate X has a continuous density with the $2d$th moment being finite. To balance the first d moments of X, we choose $\phi(X) = (X, X^2, \ldots, X^d)$. Let $Z_j = \sum_{i=1}^{d} X_i^j$, $j = 1, \ldots, d$, where $X_1, \ldots, X_d$ are independent copies of X. Then there is an open set O on which the density of $(X_1, \ldots, X_d)$ is positive and $(X_1, \ldots, X_d) \rightarrow (Z_1, \ldots, Z_d)$ is a one to one map from O to an open set $\widetilde{O}$. Then $(Z_1, \ldots, Z_d)$ has a continuous positive density on $\widetilde{O}$. That is, the dth convolution of the distribution of $\phi(X)$ has a continuous positive density on $\widetilde{O}$. Hence, Assumptions 2 (with $\gamma = 2$) and 4 are satisfied. The generalization to the multivariate case is straightforward.

4. Treatment Effect Estimation

4.1. Additive Treatment Effect

We use the Rubin potential outcomes model (Rubin 1974) to define the treatment effect. For each unit i, denote by $Y_i(1)$ and $Y_i(0)$ the potential outcomes under the treatment and control, respectively. The observed outcome is $Y_i = T_iY_i(1) + (1 - T_i)Y_i(0)$. The treatment effect is defined as $\tau = E\{Y_i(1)\} - E\{Y_i(0)\}$. To estimate τ, we consider the estimator based on the difference in observed sample means

$$\widehat{\tau} = \bar{Y}_1 - \bar{Y}_0 = \frac{\sum_{i=1}^{n} T_iY_i}{n_1} - \frac{\sum_{i=1}^{n}(1-T_i)Y_i}{n_0}.$$

Consider the data generating model in which the treatment effect is constant for all units,

$$Y_i(a) = \mu_a + \beta^{\mathrm{T}}\phi(X_i) + \varepsilon_i, \quad a = 0, 1, \tag{3}$$

where μ_1 and μ_0 are the main effects of the treatment and control, respectively, and $\phi(X_i)$ is the q-dimensional covariate features defined previously. The random error ε_i is independent and identically distributed with zero mean and a finite variance σ_ε^2 and is independent of the covariates.

Example 4.1 (First-order linear model). If $\phi(X_i) = X_i$, model (3) reduces to a typical linear model with only first-order covariate effects,

$$Y_i(a) = \mu_a + \sum_{j=1}^{p}\beta_j x_{ij} + \varepsilon_i, \quad a = 0, 1, \tag{4}$$

where β_j are covariate coefficients.

Example 4.2 (Second-order linear model). Another commonly used model, which is a special case of model (3), is a linear model with covariate effects up to the second-order,

$$Y_i(a) = \mu_a + \sum_{j=1}^{p} \beta_j x_{ij} + \sum_{j=1}^{p}\sum_{j' \le j} \beta_{jj'} x_{ij} x_{ij'} + \varepsilon_i, \quad a = 0, 1, (5)$$

where $\beta_{jj'}$ are coefficients of the quadratic terms, if $j = j'$, and the interactions, if $j' < j$.

If there are approximately the same numbers of units in each treatment, such that $n_1 - n_0 = o_P(\sqrt{n})$, it is shown in the proof of Theorem 4.1 in the Supplementary Appendix that

$$\sqrt{n}(\widehat{\tau} - \tau) = \frac{2}{\sqrt{n}}\left\{\sum_{i=1}^{n}(2T_i - 1)\beta^{\mathrm{T}}\phi(X_i) + \sum_{i=1}^{n}(2T_i - 1)\varepsilon_i\right\} + o_P(1).$$

It can be seen that the variability of $\widehat{\tau}$ consists of two components, one from the covariates and one from the random errors. Therefore, one goal of covariate balancing is to increase estimation precision by eliminating the variability due to the covariates. The best case is that the asymptotic variance of $\widehat{\tau}$ is attributable only to the random errors, that is $4\sigma_\varepsilon^2$.

Definition 1 (Optimal precision). Given the data generating model (3), we state that $\widehat{\tau}$ achieves optimal precision under a randomization procedure if $\sqrt{n}(\widehat{\tau} - \tau) \xrightarrow{D} N(0, 4\sigma_\varepsilon^2)$.

Remark 4.1. Under the first-order linear model (4), it has been shown in the literature that optimal precision can be achieved by a variety of CAR procedures, such as stratified biased coin design, Pocock and Simon's minimization, the method of Hu and Hu (2012) for discrete covariates (Shao, Yu, and Zhong 2010; Ma, Hu, and Zhang 2015), and the method of Qin et al. (2018) for continuous covariates (Ma et al. 2020a).

Theorem 4.1. Suppose $\sum_{i=1}^n (2T_i - 1) = o_P(\sqrt{n})$ and $\sum_{i=1}^n (2T_i - 1)\phi(X_i) = o_P(\sqrt{n})$ hold under the proposed randomization procedure. Then the estimated treatment effect $\widehat{\tau}$ achieves optimal precision under the data generating model (3), that is, $\sqrt{n}(\widehat{\tau} - \tau) \xrightarrow{D} N(0, 4\sigma_\varepsilon^2)$.

The assumption on the balancing of covariate features $\phi(X_i)$, that is, $\sum_{i=1}^n (2T_i - 1)\phi(X_i) = o_P(\sqrt{n})$, is satisfied by Theorem 3.1. Also, a constant function can be added into the feature map to ensure $\sum_{i=1}^n (2T_i - 1) = o_P(\sqrt{n})$. The next corollary corresponds to the second-order linear model (5) and follows directly from Theorem 4.1 and Corollary 3.2.

Corollary 4.2. Suppose the same assumptions hold as in Corollary 3.2. Then the estimated treatment effect $\widehat{\tau}$ achieves optimal precision under the COV procedure with $w_0 > 0, w_1 > 0$, and $w_2 > 0$ and the data generating model (5), that is, $\sqrt{n}(\widehat{\tau} - \tau) \xrightarrow{D} N(0, 4\sigma_\varepsilon^2)$.

We provide a consistent estimator for σ_ε^2 so that valid inference can be drawn based on Theorem 4.1. Consider the following regression:

$$Y_i = \mu_1 T_i + \mu_0(1 - T_i) + \beta^{\mathrm{T}}\phi(X_i) + \varepsilon_i. \qquad (6)$$

Let $(\widehat{\mu}_1, \widehat{\mu}_0, \widehat{\beta})$ be the ordinary-least-squares (OLS) estimator for (μ_1, μ_0, β). Then, let $\widehat{\sigma}_\varepsilon^2$ be the OLS estimator for the error variance, that is, $\widehat{\sigma}_\varepsilon^2 = (n - q - 2)^{-1}\sum_{i=1}^n \widehat{\varepsilon}_i^2$, where $\widehat{\varepsilon}_i = Y_i - \{\widehat{\mu}_1 T_i + \widehat{\mu}_0(1 - T_i) + \widehat{\beta}^{\mathrm{T}}\phi(X_i)\}$. The following theorem establishes the consistency of $\widehat{\sigma}_\varepsilon^2$. The coverage properties of the confidence intervals constructed using $\widehat{\sigma}_\varepsilon^2$ are assessed by simulations in Section B.5 in the Supplementary Appendix.

Theorem 4.3. Suppose $\sum_{i=1}^n (2T_i - 1) = o_P(\sqrt{n})$ and $\sum_{i=1}^n (2T_i - 1)\phi(X_i) = o_P(\sqrt{n})$ hold under the proposed randomization procedure. Then $\widehat{\sigma}_\varepsilon^2$ is a consistent estimator for σ_ε^2 under the data generating model (3), that is, $\widehat{\sigma}_\varepsilon^2 \xrightarrow{P} \sigma_\varepsilon^2$.

4.2. General Outcome Model

We introduce additional notation and assumptions to study treatment effect estimation under a more general model for potential outcomes. Denote the centered conditional expectations of potential outcomes by

$$m_a(X_i) = E\{Y_i(a)|X_i\} - E\{Y_i(a)\}$$

and the error terms by

$$\varepsilon_{ia} = Y_i(a) - E\{Y_i(a)|X_i\}$$

for $a = 0, 1$. We observe that $Y_i(a) - E\{Y_i(a)\} = m_a(X_i) + \varepsilon_{ia}$, with $E\{m_a(X_i)\} = 0$ and $E(\varepsilon_{ia}|X_i) = E(\varepsilon_{ia}) = 0, a = 0, 1$.

Assumption 5. $\{Y_i(1), Y_i(0), X_i\}_{i=1}^n$ are independent copies of $\{Y(1), Y(0), X\}$. Moreover, $E[\mathrm{var}\{Y_i(a)|X_i\}] > 0$ and $E\{Y_i^2(a)\} < \infty, a = 0, 1$.

The assumption $E[\mathrm{var}\{Y_i(a)|X_i\}] > 0$ is made to exclude degenerate situations, and $E\{Y_i^2(a)\} < \infty$ allows the use of certain laws of large numbers and central limit theorems, including Theorems 3.5 and 3.6. Similar assumptions have been used for inference under CAR-type procedures. See, for example, Bugni, Canay, and Shaikh (2018) and Bai, Romano, and Shaikh (in press).

Under the assumptions that ensure that $\{\Lambda_n\}$ is ψ-irreducible, we prove that $\widehat{\tau}$ is asymptotically normal. By Theorem 3.6, we have

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n}(2T_i - 1)m(X_i) \xrightarrow{D} N(0, \sigma_m^2),$$

$$m(X_i) = \frac{m_1(X_i) + m_0(X_i)}{2},$$

where, with slight notation abuse, we denote by $m(X_i)$ the average of two centered conditional expectations of potential outcomes and by $\sigma_m^2 \ge 0$ the asymptotic variance of $m(X_i)$.

Theorem 4.4. Suppose Assumptions 1, 2 (with $\gamma > 5$), 4 and 5 hold. Then $\sqrt{n}(\widehat{\tau} - \tau) \xrightarrow{D} N(0, \sigma^2)$, with $\sigma^2 = 2\mathrm{var}(\varepsilon_{i1}) + 2\mathrm{var}(\varepsilon_{i0}) + \mathrm{var}\{m_1(X_i) - m_0(X_i)\} + 4\sigma_m^2$.

Note that the terms $2\{\mathrm{var}(\varepsilon_{i1}) + \mathrm{var}(\varepsilon_{i0})\}$ and $\mathrm{var}\{m_1(X_i) - m_0(X_i)\}$ are invariant with respect to the specific feature maps used by the proposed CAR procedures. Then the randomization specific component of the asymptotic variance of $\widehat{\tau}$ is determined by

$$\sum_{i=1}^{n}(2T_i-1)m(X_i)=\left\{\sum_{i:T_i=1}m(X_i)-\sum_{i:T_i=0}m(X_i)\right\}. \tag{7}$$

Hence, Theorem 4.4 explicitly reveals the relationship between the precision of the difference-in-means estimator $\widehat{\tau}$ and the imbalance in the conditional expectation functions.

The feature map $\phi(X)$ is ideally chosen to minimize the variability of $\sum_{i=1}^{n}(2T_i-1)m(X_i)$ for higher precision. In particular, if we impose some restrictions on the functional form of $m(X)$, then $\phi(X)$ can be constructed so that $m(X) \in \operatorname{span}\{\phi(X)\}$. In this case, $\sigma_m^2 = 0$ and the corresponding $\widehat{\tau}$ has a possibly minimum asymptotic variance. In general, the functional form of $m(X)$ may be unknown. However, because the conditional expectation functions are usually nonlinear, the span of the higher q-dimensional $\phi(X)$ is more likely to contain or well approximate $m(X)$ than that of the original lower p-dimensional X. These arguments further justify the need to balance general covariate features.

Remark 4.2. The results in Theorem 4.4 provide some insights on the use of the KER procedure when $m(X)$ is a continuous function, as the Gaussian kernel is universal: the basis functions of the Gaussian kernel can approximate any continuous function.

Remark 4.3. When all of the baseline covariates are available before randomization, Kallus (2018) derived the same imbalance metric as that in (7) by considering the minimax variance among all possible allocations. Integer programming algorithms were then used to obtain the optimal allocation. However, these algorithms do not apply to settings in which the units enter the experiment over a period of time, such as in sequential clinical trials. In contrast, our proposed CAR procedures adaptively assign the treatments and are thus more feasible for sequential clinical trials.

Remark 4.4. Theorem 4.4 provides a basis for hypothesis testing and constructing confidence intervals for the treatment effect if a consistent variance estimator is available. However, compared with the discrete case, a sample analog variance estimator is more difficult to obtain (Bugni, Canay, and Shaikh 2018, 2019; Ye, Yi, and Shao 2020; Ma, Tu, and Liu 2020b), because σ_m^2 does not have a closed form. Alternatively, it is likely that bootstrap methods could be used to draw valid inference under the proposed procedures (Shao, Yu, and Zhong 2010; Ma, Hu, and Zhang 2015; Zhang and Zheng 2020); but this remains a conjecture.

5. Simulation Studies

5.1. Convergence Rates of Imbalance Vectors

We first evaluate the convergence rates of imbalance vectors under different randomization procedures, including complete randomization (CR), the proposed COV procedures with different weights, and KER. The covariates X_i are simulated from multivariate normal distributions $N(0, I_p)$ with $p = 1, 2$ and 4, where I_p is the p-dimensional identity matrix. The sample sizes are $n = 200, 500, 800$, and 2000. For each randomization procedure, the imbalances at different levels are investigated, including the differences in numbers of units $\sum(2T_i - 1) = n_1 - n_0$, the first- and second-moment covariate imbalances $\sum(2T_i-1)x_{i,1}$ and $\sum(2T_i-1)x_{i,1}^2$, and the covariate imbalance measured by $\sum(2T_i-1)\exp(-||X_i||^2)$. The standard deviations of these imbalance vectors are used to evaluate the convergence rate and are given in Table 1. Due to symmetry, only the results of the first dimensions of the covariates, $x_{i,1}$, are reported. All of the simulations in this and subsequent sections are based on 5000 replicates. Additional simulations under different covariate assumptions and comparisons with other CAR procedures are presented in Section B.3 in the Supplementary Appendix.

Under CR, COV with $w_0 = 0$, and KER, the imbalances $\sum(2T_i-1) = n_1 - n_0$ become more variable as the sample size increases. The standard deviations increase with approximately $O_P(\sqrt{n})$ rates under both CR and COV with $w_0 = 0$, and with faster rates than $O_P(\sqrt{n})$ under KER. In contrast, under the two COV procedures with $w_0 = 1$, the imbalances tend to stabilize, meaning that they do not increase as the sample size increases. These results suggest that if the numbers of units in each treatment need balancing, a positive weight of w_0 should be imposed.

Second, the standard deviations of $\sum(2T_i - 1)x_{i,1}$ stabilize under all four COV procedures, regardless of whether $w_0 = 0$. The cases in which $w_2 = 0$ perform better than those in which $w_2 > 0$. This result is expected because the first two COV procedures only balance the first moment of the covariates. Under KER, the standard deviations become larger as the sample size increases, but the rates are faster than the $O_P(\sqrt{n})$ rates observed under CR.

For the second-moment imbalances $\sum(2T_i - 1)x_{i,1}^2$, the two COV procedures with $w_2 > 0$ are the only procedures that ensure stabilization of the standard deviations. Under both CR and COV with $w_2 = 0$, the standard deviations increase with approximately $O_P(\sqrt{n})$ rates. Under KER, the standard deviations increase at rates faster than $O_P(\sqrt{n})$, but they do not stabilize. In summary, COV ($w_2 > 0$) is the best performing procedure for covariate balancing when taking into account both the first and second moments.

Finally, for the imbalance measured by $\sum(2T_i - 1)\exp(-||X_i||^2)$, the KER procedure has the smallest standard deviations and outperforms all other methods. The standard deviations tend to stabilize under KER as the sample size increases. Note that $\exp(-||X_i||^2)$ is the first basis function of the Gaussian kernel. The imbalances on subsequent basis functions are presented in Table B.1 in Section B.1 in the Supplementary Appendix. In these additional simulation results, the imbalances stabilize on the first few basis functions, but increase with sample size for the later basis functions. See Section B.1 in the Supplementary Appendix for further discussion of these results.

5.2. Imbalance Measures under COV

In this section, we evaluate imbalance measures based on the first two moments under COV. We simulate the covariate X_i again according to multivariate normal distributions $N(0, I_p)$.

Table 2. Means and standard deviations (STD) of imbalance measures based on covariate means and covariance matrices under COV.

Randomization	n	$n^2\|\|\bar{X}_1 - \bar{X}_0\|\|^2$		$n^2\|\|S_1 - S_0\|\|_F^2$	
		Mean	STD	Mean	STD
CR	200	1662.05	1645.81	4869.66	4070.06
	500	4002.37	4022.40	11916.09	9917.10
	800	6413.39	6419.536	19248.63	15770.86
COV	200	18.36	20.52	4896.53	4114.22
$(w_0, w_1, w_2) = (1, 1, 0)$	500	18.33	20.61	12563.91	10092.17
	800	18.98	21.66	19747.08	15980.93
COV	200	41.93	49.01	246.93	255.44
$(w_0, w_1, w_2) = (1, p, 1)$	500	41.84	48.72	246.13	257.24
	800	42.52	49.26	245.03	257.36
COV	200	21.86	24.97	395.54	344.76
$(w_0, w_1, w_2) = (1, p, 1/4)$	500	22.32	26.00	392.17	355.20
	800	22.79	26.94	403.23	366.29

Table 3. Means ($n\times$variances) of treatment effect estimates under various randomization procedures and models.

Randomization	n	Model			
		1	2	3	4
CR	200	1.00(12.03)	1.00(32.47)	0.99(6.94)	1.00(16.71)
	500	1.00(12.16)	1.00(31.29)	1.00(6.90)	1.00(17.39)
	800	1.00(11.71)	1.00(31.93)	1.00(6.82)	1.00(16.69)
COV	200	1.00(4.13)	1.01(25.97)	1.00(6.46)	1.00(9.13)
$(w_0, w_1, w_2) = (1, 1, 0)$	500	1.00(4.11)	0.99(24.54)	1.00(6.57)	1.00(9.15)
	800	1.00(4.08)	1.00(24.76)	1.00(6.85)	1.00(9.01)
COV	200	1.00(4.31)	1.00(5.48)	1.00(6.28)	1.00(4.68)
$(w_0, w_1, w_2) = (1, p, 1)$	500	1.00(4.15)	1.00(4.54)	1.00(6.40)	1.00(4.64)
	800	1.00(4.14)	1.00(4.28)	1.00(6.57)	1.00(4.59)
KER	200	1.00(4.94)	1.00(9.73)	1.01(4.08)	0.99(5.67)
	500	1.00(4.45)	1.00(7.63)	0.99(4.02)	1.00(4.98)
	800	1.00(4.37)	1.00(6.92)	1.00(3.91)	1.00(4.76)

The proposed COV procedure is used to assign the units to treatment groups.

In view of the results in Section 5.1, we set $w_0 = 1$ to improve the similarity between the numbers of units in each treatment. Three different sets of weights are considered, $(w_0, w_1, w_2) = (1, 1, 0), (1, p, 1)$ and $(1, p, 1/4)$. After all of the units are assigned, the imbalances in covariate means and covariance matrices, measured by $||\bar{X}_1 - \bar{X}_0||^2$ and $||S_1 - S_0||_F^2$, are recorded. CR is also applied to these simulations for comparison purposes. Three different sample sizes, $n = 200, 500$, and 800 are used in the simulations. Table 2 shows the means and standard deviations of $n^2||S_1 - S_0||_F^2$ and $n^2||\bar{X}_1 - \bar{X}_0||^2$ under different randomization procedures. For simplicity, only the results of $p = 2$ are listed. The results of $p = 1$ and $p = 4$ have similar patterns and are therefore omitted. The histograms of the imbalance measures are plotted in Figures B.1, B.2, and B.3 in Section B.2 in the Supplementary Appendix.

From Table 2,the means and standard deviations of both $n^2||\bar{X}_1 - \bar{X}_0||^2$ and $n^2||S_1 - S_0||_F^2$ are stable as the sample size increases under COV with $w_2 > 0$. These results hold for both sets of weights, indicating that the inclusion of $w_2 > 0$ is more important than the value of w_2 in controlling the convergence rate. However, compared with the case with weights $(w_0, w_1, w_2) = (1, p, 1)$, the procedure with $(w_0, w_1, w_2) = (1, p, 1/4)$ is better for balancing means, but worse for balancing covariance matrices because it imposes more weight on the first moment in the overall imbalance measure.

In contrast, the other two procedures, COV with $w_2 = 0$ and CR, do not control the covariance matrix imbalances well. Both the means and the standard deviations of $n^2||S_1 - S_0||_F^2$ increase as the sample size increases, growing at approximately the same rates as $O_P(n)$. Although COV with $w_2 = 0$ shows the smallest imbalances of covariate means among the procedures, its covariance matrix imbalances are similar to those for CR, making $w_2 = 0$ a less satisfactory choice for COV if balancing of covariances is deemed critical.

5.3. Treatment Effect Estimation

We simulate two-dimensional covariates $X_i \sim N(0, I_2)$ and apply a CAR procedure to obtain the treatment assignments T_i. The observed outcome is $Y_i = T_i Y_i(1) + (1 - T_i) Y_i(0)$. For $a = 0, 1$ and $i = 1, \ldots, n$, the potential outcomes $Y_i(a), a = 0, 1$, are simulated according to the following data generating models:

Model 1: $Y_i(a) = \mu_a + x_{i1} + x_{i2} + \varepsilon_i$,
Model 2: $Y_i(a) = \mu_a + x_{i1} + x_{i2} + x_{i1}^2 + x_{i2}^2 + x_{i1}x_{i2} + \varepsilon_i$,
Model 3: $Y_i(a) = \mu_a + 2(1 + x_{i1} + x_{i2} + x_{i1}x_{i2}) \exp(-x_{i1}^2 - x_{i2}^2) + \varepsilon_i$,
Model 4: $Y_i(a) = \mu_a + x_{i1} + x_{i2} + x_{i1}x_{i2} + \exp(-x_{i1}^2) + \exp(-x_{i2}^2) + \varepsilon_i$,

where $\mu_1 = 1$, $\mu_0 = 0$, and $\varepsilon_i \sim N(0, 1)$ is independent of the covariates.

Model 1 represents the simplest case, in which covariate effects are both linear and additive. In Model 2, the outcome variable follows a second-order linear model, which contains both quadratic and interaction effects of the covariates. Model 3 includes the first four basis functions induced by the two-dimensional Gaussian kernel, which are $\exp(-x_{i1}^2 - x_{i2}^2)$, $x_{i1} \exp(-x_{i1}^2 - x_{i2}^2)$, $x_{i2} \exp(-x_{i1}^2 - x_{i2}^2)$ and $x_{i1}x_{i2} \exp(-x_{i1}^2 - x_{i2}^2)$, respectively. In Model 4, the covariates exhibit additive but nonlinear effects on the outcomes, which include the first-order and interaction effects of the covariates and the two basis functions associated with the two-dimensional Gaussian kernel. For other and more general model settings, such as nonadditive treatment effects and error terms that are correlated with covariates or not normally distributed, additional simulations are conducted in Section B.3.3 in the Supplementary Appendix. We also present comparisons with other CAR procedures.

To assess the estimated treatment effect $\widehat{\tau} = \bar{Y}_1 - \bar{Y}_0$, CR and three CAR procedures are compared, including KER and two COV procedures with different weights. The sample sizes are $n = 200, 500$, and 800. Table 3 presents the means and variances of the treatment effect estimates under different data generating models and randomization procedures.

As seen from Table 3, the means of the treatment effect estimates $\widehat{\tau}$ are equal or very close to the true value, indicating the asymptotic unbiasedness of $\widehat{\tau}$ for all of the randomization procedures under consideration. We then compare the variances of these estimators as indicators of estimation precision.

For Model 1, with only first-order covariate effects, both COV procedures achieve optimal precision, meaning that the variances are approximately four times the variances of the random errors. This finding is consistent with the theoretical results in Section 4.1. The variances under KER are slightly larger than those under COV because KER does not directly balance the first-order covariate effects. For Model 2, the COV procedures with $w_2 > 0$ exhibit the smallest variances, closely followed by KER. Both KER and COV with $w_2 > 0$ have much better performance than CR and COV with $w_2 = 0$. For this model, the inferior performance of COV with $w_2 = 0$ is expected because it is only balances the first moment of the covariates.

Model 3 includes the first few basis functions of the Gaussian kernel. KER shows the minimum variances among the three methods, and it appears that these variances are solely attributable to the random errors. Finally, the functional form of the covariate effects in Model 4 is continuous but does not belong to the span of COV or KER basis functions. In this scenario, KER and COV with $w_2 > 0$ perform similarly well, but CR and COV with $w_2 = 0$ are less effective.

In summary, the performances of KER and COV with $w_2 > 0$ are relatively robust under all of the scenarios studied, with each having its own advantages. In contrast, although it achieves optimal precision under the first-order linear model, COV with $w_2 = 0$ is generally less competitive under more complex nonlinear models. This finding demonstrates the limitations presented by balancing only the first moment of covariates in randomization.

6. Clinical Trial Example

We present a clinical trial example to illustrate the covariate balance and estimation precision advantages of the proposed CAR procedures. The example is based on a randomized clinical trial to compare nefazodone, the cognitive behavioral-analysis system of psychotherapy, and their combination for the treatment of chronic depression (Keller et al. 2000). We consider only the data of nefazodone and the combination treatment because this article focuses on the scenario of two treatments. The outcome variable is the last observed 24-item Hamilton Rating Scale for Depression post treatment (FinalHAMD). The covariates of interest are age and the 24-item Hamilton Rating Scale for Depression (HAMD24) at baseline. There were no significant differences between the two treatment groups with respect to the covariates, although the age variances were dissimilar (122 vs. 105).

We analyzed the data with an additive model that included both age and HAMD24 at baseline as covariates,

$$\begin{aligned}\text{FinalHAMD}_i = \mu_1 T_i + \mu_0(1 - T_i) + f_1(\text{AGE}_i) \\ + f_2(\text{HAMD24}_i) + \varepsilon_i,\end{aligned} \tag{8}$$

where $\varepsilon_i \sim N(0, \sigma_a^2)$. The treatment effect of the combination treatment ($T_i = 1$) over nefazodone ($T_i = 0$) is $\tau = \mu_1 - \mu_0$. The estimated smooth effect of the covariates are plotted in Figure 1. As can be seen from the figure, the age effect appears quadratic, whereas the effect of baseline HAMD24 is more likely to be linear. Hence, we also fitted a quadratic model for the data,

$$\begin{aligned}\text{FinalHAMD}_i = \mu_1 T_i + \mu_0(1 - T_i) + \beta_1 \text{AGE}_i + \beta_2 \text{AGE}_i^2 \\ + \beta_3 \text{HAMD24}_i + \varepsilon_i,\end{aligned} \tag{9}$$

where $\varepsilon_i \sim N(0, \sigma_q^2)$. The covariate effects, including the quadratic age effect, are all significant in this model.

For comparison, four randomization procedures, including CR, two COV procedures with $w_2 = 0$ (balancing covariate means only) and $w_2 > 0$ (balancing both covariate means and covariance matrices), and KER, were applied to reassign the patients receiving nefazodone or the combination treatment. To minimize the impact of variability differences between the two covariates, we used standardized values in the three CAR procedures. After assignment of the patients, outcomes were generated with the estimated parameters from the additive model (8) and the quadratic model (9) as the true values. Specifically, the true values of τ were set to -4.95 and -4.87 for the additive model and the quadratic model, respectively. It was also noted that the variability of the random errors dominated that of the covariates in both models, so we set $\sigma_a = \sigma_q = 2$ to make the effects of random errors and covariates more comparable. We further estimated the treatment effect by the differences in observed sample means $\widehat{\tau}$ as described in Section 4. Each

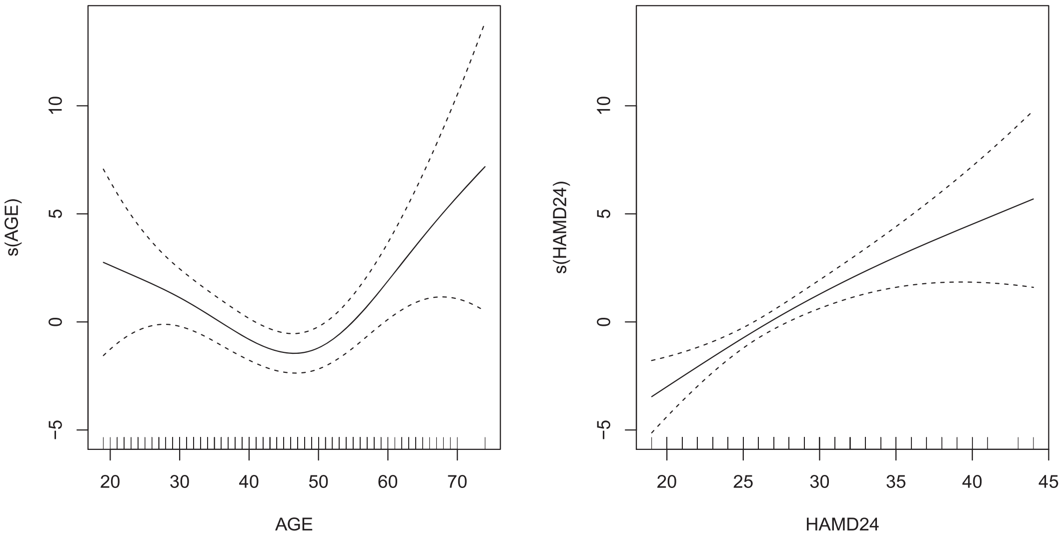

Figure 1. Estimated smooth effects of the covariates under the additive model for the depression data.

Table 4. Absolute differences in covariate means and variances between treatments under various randomization procedures for the depression data.

Randomization	Absolute difference in means		Absolute difference in variances	
	Age	HAMD24	Age	HAMD24
CR	0.81	0.38	10.37	2.53
COV $(w_0, w_1, w_2) = (1, 1, 0)$	0.06	0.02	9.76	2.06
COV $(w_0, w_1, w_2) = (1, p, 1)$	0.07	0.03	1.52	0.31
KER	0.17	0.09	3.59	1.17

Table 5. Means and standard deviations (STD) of treatment effect estimates under various randomization procedures and models for the depression data.

Randomization	Additive model $(\tau = -4.95)$		Quadratic model $(\tau = -4.87)$	
	Mean	STD	Mean	STD
CR	−4.95	0.29	−4.87	0.30
COV $(w_0, w_1, w_2) = (1, 1, 0)$	−4.95	0.23	−4.87	0.23
COV $(w_0, w_1, w_2) = (1, p, 1)$	−4.95	0.19	−4.87	0.19
KER	−4.95	0.20	−4.87	0.20

scenario was simulated 5000 times. To evaluate each covariate balance, we calculated the average of the absolute differences in covariate means and variances between treatments. The results are listed in Table 4. The estimation precision was assessed using simulated means and standard deviations of the treatment effect estimates $\widehat{\tau}$. The results are presented in Table 5.

Table 4 shows that better covariate balance can be obtained by the proposed procedures, COV and KER, than by CR. COV with $w_2 = 0$ provides the best balance of covariate means, but the variance differences are similar to those from CR, making the overall performance of COV with $w_2 = 0$ less satisfactory than the other two CAR procedures. COV with $w_2 > 0$ outperforms all of the other methods in reducing the absolute differences in covariate variances and is only slightly worse than COV with $w_2 = 0$ in terms of balancing means. The performance of KER falls between CR and COV with $w_2 > 0$ because it does not directly balance the first and second moments of the covariates.

From Table 5 it can also be seen that the variances of the treatment effect estimates from the three CAR procedures are all lower than the corresponding values from CR, indicating that covariate balancing improved the estimation precision. Because of the apparent quadratic effect in the data, the precision improvement is most noticeable for COV with $w_2 > 0$ under both the additive and quadratic models.

7. Conclusion

In this article, we propose a new and unified family of CAR procedures that balance general covariate features. The convergence rates of imbalance vectors and the treatment effect estimation are evaluated both theoretically and by numerical studies. It is shown that the procedures have favorable theoretical and practical properties relative to complete randomization and existing CAR procedures. The results of this article could be further generalized in several ways. First, it is important to extend the framework to handle multiple treatments that may have unequal target allocations (Bugni, Canay, and Shaikh 2019; Hu and Zhang 2020). Second, other kernel-induced imbalance measures could be assessed for application within the proposed framework. Finally, adjustment for additional baseline covariates could be considered to improve precision (Ma, Tu, and Liu 2020b; Ye, Yi, and Shao 2020).

Supplementary Materials

The Supplementary Materials include (1) the Appendix containing technical proofs and additional simulation results and (2) the code used for this article.

Acknowledgments

The authors thank the Co-Editor, the Associate Editor, and the two anonymous referees for many helpful comments and suggestions. The authors also thank Yujia Gu and Xiao Liu for excellent research assistance.

Funding

Wei Ma's research was supported by the National Key R&D Program of China (grant no. 2018YFC2000302) and the National Natural Science Foundation of China (grant nos. 12171476, 11731011). Li-Xin Zhang's research was supported by the National Natural Science Foundation of China (grant nos. 11731012, 12031005), Ten Thousands Talents Plan of Zhejiang Province (grant no. 2018R52042) and the Fundamental Research Funds for the Central Universities. Ping Li's work was conducted at Baidu Research, USA. Part of Feifang Hu's research was conducted while he was visiting Baidu Research, USA, as a technical advisor to the company. He thanks Baidu Research, USA, for their generous support.

References

Aronszajn, N. (1950), "Theory of Reproducing Kernels," *Transactions of the American Mathematical Society*, 68, 337–404. [4]

Atkinson, A. C. (1982), "Optimum Biased Coin Designs for Sequential Clinical Trials with Prognostic Factors," *Biometrika*, 69, 61–67. [1,5]

——— (2002), "The Comparison of Designs for Sequential Clinical Trials with Covariate Information," *Journal of the Royal Statistical Society*, Series A, 165, 349–373. [1]

Bai, Y., Romano, J. P., and Shaikh, A. M. (in press), "Inference in Experiments with Matched Pairs," *Journal of the American Statistical Association*, DOI: 10.1080/01621459.2021.1883437. [7]

Baldi Antognini, A., and Zagoraiou, M. (2011), "The Covariate-Adaptive Biased Coin Design for Balancing Clinical Trials in the Presence of Prognostic Factors," *Biometrika*, 98, 519–535. [1,2]

Begg, C. B., and Iglewicz, B. (1980), "A Treatment Allocation Procedure for Sequential Clinical Trials," *Biometrics*, 36, 81–90. [1]

Begg, C. B., and Kalish, L. A. (1984), "Treatment Allocation for Nonlinear Models in Clinical Trials: The Logistic Model," *Biometrics*, 40, 409–420. [1]

Bugni, F. A., Canay, I. A., and Shaikh, A. M. (2018), "Inference under Covariate-Adaptive Randomization," *Journal of the American Statistical Association*, 113, 1784–1796. [2,7,8]

——— (2019), "Inference under Covariate-Adaptive Randomization with Multiple Treatments," *Quantitative Economics*, 10, 1747–1785. [8,11]

Efron, B. (1971), "Forcing a Sequential Experiment to be Balanced," *Biometrika*, 58, 403–417. [3]

Endo, A., Nagatani, F., Hamada, C., and Yoshimura, I. (2006), "Minimization Method for Balancing Continuous Prognostic Variables between Treatment and Control Groups using Kullback-Leibler Divergence," *Contemporary Clinical Trials*, 27, 420–431. [1]

Frane, J. W. (1998), "A Method of Biased Coin Randomization, its Implementation, and its Validation," *Drug Information Journal*, 32, 423–432. [1]

Gretton, A., Borgwardt, K., Rasch, M. J., Scholkopf, B., and Smola, A. J. (2008), "A Kernel Method for the Two-Sample Problem," arXiv preprint arXiv:0805.2368. [4]

Hoehler, F. K. (1987), "Balancing Allocation of Subjects in Biomedical Research: A Minimization Strategy based on Ranks," *Computers and Biomedical Research*, 20, 209–213. [1]

Hofmann, T., Schölkopf, B., and Smola, A. J. (2008), "Kernel Methods in Machine Learning" *The Annals of Statistics*, 36, 1171–1220. [4]

Hu, F., Hu, Y., Ma, Z., and Rosenberger, W. F. (2014), "Adaptive Randomization for Balancing over Covariates," *Wiley Interdisciplinary Reviews: Computational Statistics*, 6, 288–303. [1]

Hu, F., and Zhang, L.-X. (2020), "On the Theory of Covariate-Adaptive Designs," arXiv preprint arXiv:2004.0299. [3,11]

Hu, Y., and Hu, F. (2012), "Asymptotic Properties of Covariate-Adaptive Randomization," *The Annals of Statistics*, 40, 1794–1815. [1,2,3,5,7]

Jiang, F., Ma, Y., and Yin, G. (2018), "Kernel-based Adaptive Randomization toward Balance in Continuous and Discrete Covariates," *Statistica Sinica*, 28, 2841–2856. [1,4,5]

Kallus, N. (2018), "Optimal a priori Balance in the Design of Controlled Experiments," *Journal of the Royal Statistical Society*, Series B, 80, 85–112. [8]

Keller, M. B., McCullough, J. P., Klein, D. N., Arnow, B., Dunner, D. L., Gelenberg, A. J., Markowitz, J. C., Nemeroff, C. B., Russell, J. M., Thase, M. E., Trivedi, M. H., and Zajecka, J. (2000), "A Comparison of Nefazodone, the Cognitive Behavioral-Analysis System of Psychotherapy, and their Combination for the Treatment of Chronic Depression," *New England Journal of Medicine*, 342, 1462–1470. [2,10]

Li, X., Zhou, J., and Hu, F. (2019), "Testing Hypotheses under Adaptive Randomization with Continuous Covariates in Clinical Trials," *Statistical Methods in Medical Research*, 28, 1609–1621. [1,2,3,5]

Lin, Y., and Su, Z. (2012), "Balancing Continuous and Categorical Baseline Covariates in Sequential Clinical Trials using the Area between Empirical Cumulative Distribution Functions," *Statistics in Medicine*, 31, 1961–1971. [1]

Lin, Y., Zhu, M., and Su, Z. (2015), "The Pursuit of Balance: An Overview of Covariate-Adaptive Randomization Techniques in Clinical Trials," *Contemporary Clinical Trials*, 45, 21–25. [1]

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. [2,7,8]

Ma, W., Qin, Y., Li, Y., and Hu, F. (2020a), "Statistical Inference for Covariate-Adaptive Randomization Procedures," *Journal of the American Statistical Association*, 115, 1488–1497. [2,7]

Ma, W., Tu, F., and Liu, H. (2020b), "Regression Analysis for Covariate-Adaptive Randomization: A Robust and Efficient Inference Perspective," arXiv preprint arXiv:2009.02287. [2,8,11]

Ma, Z., and Hu, F. (2013), "Balancing Continuous Covariates based on Kernel Densities," *Contemporary Clinical Trials*, 34, 262–269. [1,4]

McEntegart, D. J. (2003), "The Pursuit of Balance using Stratified and Dynamic Randomization Techniques: An Overview," *Therapeutic Innovation & Regulatory Science*, 37, 293–308. [1]

Meyn, S. P., and Tweedie, R. L. (2009), *Markov Chains and Stochastic Stability* (2nd ed.), Cambridge: Cambridge University Press. [2,5,6]

Morgan, K. L., and Rubin, D. B. (2012), "Rerandomization to Improve Covariate Balance in Experiments," *The Annals of Statistics*, 40, 1263–1282. [3]

Muandet, K., Fukumizu, K., Sriperumbudur, B., and Schölkopf, B. (2016), "Kernel Mean Embedding of Distributions: A Review and Beyond," arXiv preprint arXiv:1605.09522. [4]

Nishi, T., and Takaichi, A. (2004), "An Extended Minimization Method to Assure Similar Means of Continuous Prognostic Variables between Treatment Groups," *Japanese Journal of Biometrics*, 24, 43–55. [1,4]

Pocock, S. J., and Simon, R. (1975), "Sequential Treatment Assignment with Balancing for Prognostic Factors in the Controlled Clinical Trial," *Biometrics*, 31, 103–115. [1,3]

Qin, Y., Li, Y., Ma, W., and Hu, F. (2018), "Pairwise Sequential Randomization and its Properties," arXiv preprint arXiv:1611.02802v2. [1,2,3,4,5,7]

Rosenberger, W. F., and Lachin, J. M. (2015), *Randomization in Clinical Trials: Theory and Practice* (2nd ed.), Hoboken: Wiley. [1,2]

Rubin, D. B. (1974), "Estimating Causal Effects of Treatments in Randomized and Nonrandomized Studies," *Journal of Educational Psychology*, 66, 688–701. [6]

Shao, J., Yu, X., and Zhong, B. (2010), "A Theory for Testing Hypotheses under Covariate-Adaptive Randomization," *Biometrika*, 97, 347–360. [2,3,7,8]

Smith, R. L. (1984), "Sequential Treatment Allocation using Biased Coin Designs," *Journal of the Royal Statistical Society*, Series B, 46, 519–543. [1]

Sriperumbudur, B. K., Fukumizu, K., and Lanckriet, G. R. (2011), "Universality, Characteristic Kernels and RKHS Embedding of Measures," *Journal of Machine Learning Research*, 12, 2389–2410. [4]

Steinwart, I., Hush, D., and Scovel, C. (2006), "An Explicit Description of the Reproducing Kernel Hilbert Spaces of Gaussian RBF Kernels," *IEEE Transactions on Information Theory*, 52, 4635–4643. [4]

Stigsby, B., and Taves, D. R. (2010), "Rank-Minimization for Balanced Assignment of Subjects in Clinical Trials," *Contemporary Clinical Trials*, 31, 147–150. [1]

Taves, D. R. (2010), "The Use of Minimization in Clinical Trials," *Contemporary Clinical Trials*, 31, 180–184. [1]

Ye, T., Yi, Y., and Shao, J. (2020), "Inference on Average Treatment Effect under Minimization and other Covariate-Adaptive Randomization Methods," arXiv preprint arXiv:2007.09576. [8,11]

Zhang, Y., and Zheng, X. (2020), "Quantile Treatment Effects and Bootstrap Inference under Covariate-Adaptive Randomization," *Quantitative Economics*, 11, 957–982. [8]

Zhou, Q., Ernst, P. A., Morgan, K. L., Rubin, D. B., and Zhang, A. (2018), "Sequential Rerandomization," *Biometrika*, 105, 745–752. [1,3]